

**THÈSE DE DOCTORAT DE  
L'UNIVERSITÉ PIERRE ET MARIE CURIE**

Spécialité Biostatistiques

École Doctorale Pierre Louis de Santé Publique à Paris :  
Épidémiologie et Sciences de l'Information Biomédicale

Présentée par

Monsieur Thomas OBADIA

Pour obtenir le grade de

**DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE**

**L'apport des données de contacts détaillées à la  
modélisation épidémiologique : application au cas des  
infections nosocomiales à *Staphylococcus aureus***

soutenue le 15/07/2015 devant le jury composé de :

M. le Professeur Pierre-Yves BOËLLE (Université Pierre et Marie Curie)	Directeur
M. le Docteur Frédéric HAMELIN (Université de Rennes 1)	Examineur
M. le Professeur Éric MAURY (Université Pierre et Marie Curie)	Examineur
M. le Docteur Viet Chi TRAN (Université de Lille 1)	Rapporteur
Mme le Docteur Élisabeta VERGU (Institut National de la Recherche Agronomique)	Rapporteur

## **Laboratoire de rattachement**

**UMR-S 1136, INSERM – Université Pierre et Marie Curie Paris VI**

**« Institut Pierre Louis d'Épidémiologie et de Santé Publique »**

Directrice : Dominique COSTAGLIOLA

## **Équipe 1 – Surveillance et modélisation des maladies transmissibles**

Responsable : Pierre-Yves BOËLLE

Faculté de Médecine Pierre et Marie Curie, site Saint-Antoine

27, rue Chaligny

75571 Paris cedex 12

France

## Remerciements

Il me semble impossible de démarrer cette section sans commencer par remercier chaleureusement mon directeur de thèse, Pierre-Yves Boëlle. Au cours de ces 3 années, son encadrement a été sans faille et il s'est toujours montré disponible pour suggérer de nouvelles pistes à étudier ou discuter les orientations que prenait ma thèse, relire et corriger les versions successives des articles avec une assiduité rare, passer une après-midi à parser un SVG pour transformer des triangles en petits bonshommes afin de générer « la » figure... Et ce, malgré un emploi du temps obscur, que seul lui peut interpréter. Merci également de m'avoir fait confiance en me proposant une thèse presque au pied levé, dès 2011, alors que je terminais seulement mes études d'ingénieur en agronomie. Un encore plus grand « merci » pour m'avoir permis d'intégrer l'U707 à cette époque, quand les délais pour les contrats doctoraux étaient dépassés. Cette première année pré-thèse, passée à l'INSERM, m'a permis de prendre de l'avance en épidémiologie, et a sans nul doute été décisive pour terminer ma thèse dans le temps imparti.

C'est également l'occasion de remercier Alain-Jacques Valleron et Guy Thomas, qui ont facilité mon arrivée à l'U707 en 2011. Alain-Jacques Valleron m'a redirigé vers Pierre-Yves, lorsque je réfléchissais à une thèse en biostatistiques. Je n'oublie pas non plus Georges Offenstadt, premier interlocuteur de cette chaîne, qui a pris le temps de me rencontrer alors que j'étais indécis et m'a aiguillé vers Alain-Jacques Valleron. Je me souviendrai longtemps des dépannages informatiques improvisés dans le bureau de Georges !

*A bien y réfléchir, cette chaîne d'interlocuteurs successifs est un préambule amusant pour introduire une thèse qui utilise des réseaux de contacts...!*

Je tiens à remercier Viet Chi Tran et Elisabeta Vergu pour avoir accepté d'être rapporteurs de cette thèse, ainsi que Frédéric Hamelin et Éric Maury d'en être examinateurs.

Les collaborateurs (et co-auteurs) des différentes équipes impliquées dans le projet I-Bird, qui sert de base à cette thèse, ont joué un rôle crucial pour la mener à bien. Merci donc à Lulla Opatowski, Laura Temime, Judith Legrand, Anne Thiébaut, Didier Guillemot, Éric Fleury et Jean-Louis Herrmann, pour vos relectures et commentaires sur les articles et lors des réunions. Une pensée spéciale pour Lulla qui a en plus accepté de relire, dans des délais serrés, mon manuscrit de thèse !

Pour les discussions sur certaines méthodes statistiques (et les conseils sur les palettes de couleurs pour les figures !), je remercie Gilles Hejblum, ainsi que Pierre-Yves évidemment.

Ma thèse s'est toujours déroulée dans une bonne ambiance, notamment grâce à mes « co-bureaux » successifs : Stéphanie Gomme et Ségolène Charaudeau, Olivier Robineau (ça va, gros ?!), et plus récemment Julien Riou. Dans un même bureau, on sait toujours

avoir des discussions passionnantes : « dingue, j'ai réussi à imbriquer un *apply* dans un *lapply*, ça marche du premier coup et en plus ça fait ce que je veux !! » (Julien ? Olivier ?). Et dans des bureaux très proches, Romain Silhol (qui a déniché toutes les bonnes adresses pour déjeuner), Marc Carpentier, Pierre-François Busson, Bérengère Couturier, Nathanaël Lapidus. Les Boucheries et autre Petit Quercy du midi ont sans nul doute rendu ces années agréables à vos côtés !

Enfin, dans les bureaux un peu plus éloignés, je remercie Cécile Souty et Clément Turbelin, avec qui j'ai souvent échangé. La dernière partie de cette thèse contient également des éléments développés conjointement avec Eugenio Valdano, doctorant de l'unité supervisé par Vittoria Colizza : je les remercie pour l'opportunité de travailler ensemble.

Une pensée également pour les anciens du Master MSB : Élodie, Charlotte, Antoine (po-po-poooo), SLYvain, Charles, Nicolas, toujours au top pour trouver de quoi procrastiner (et pour ça, je vous dis GFYWAC) !

Parmi les amis proches, un grand merci à (M)Amandine, (mamie) Lisa, Léa, Arsène (le squatteur multirécidiviste), Fabien (stupe), Stéphane (flûte), ~~Rémi~~ Su7, Karl, Hadrien... Les sorties ensemble ont égaillé ces années, et il faudra que ça continue ! Merci à Nathalie et Marc Schneider pour les burgers savoureux du vendredi midi au Martini (« Excuse-moi, est-ce que tu pourrais fermer ta ..., s'il te plaît ? Mais j'ai dit « s'il te plaît » ?! »), en compagnie de Lisa et Amandine justement. Je n'oublie pas Cyril et Damien (VOILA-MERCI-BISOUS-!) pour le soutien psychologique/franche rigolade/mauvaise foi massive/défouloir du soir, ni Sam pour les sushis réguliers. La « team ouais-ouais » compte aussi pour beaucoup dans ce « capital bonne humeur » : Malo, GZoffrey, ~~Romain~~ Booba, Yaniv. Merci aussi à Paula pour les week-ends rigolos, et les discussions philosophiques par SMS sur l'issue de l'émission en cours... Bref, à vous tous (et à ceux que j'ai oublié dans ce paragraphe, promis, c'est pas volontaire !!), merci, car c'est aussi grâce à vous que j'ai passé 3 bonnes années de thèse.

Ensuite, il y a Patrice Clozier : le voyage formidable que tu as organisé pour Sarah et moi, à l'été 2014, m'a permis de démarrer ma dernière année de thèse en étant reposé (et – un peu- bronzé !). Résultat, quelques mois plus tard, thèse rédigée : c'est un peu grâce à toi aussi... Merci !

Évidemment, je remercie ma famille : Cécile (relectrice attentive), Édouard, Nico, et aussi Papé, Mamé, Arlette et Georges, vous y êtes « un peu » (et même « pas qu'un peu ») pour quelque chose, et j'espère que vous en êtes heureux !

Et enfin, Sarah, merci de me supporter (et des fois, ça doit pas être facile (j'ai hésité avec « merci d'avoir la chance de me supporter »... bon, OK, j'arrête)) et d'être là depuis le début de ma thèse et même plus. Je t'aime.



## Résumé de la thèse

La modélisation mathématique en épidémiologie utilise traditionnellement l'hypothèse de mélange homogène des individus dans la population. Si certains modèles structurés permettent de tenir compte de l'hétérogénéité apparente (spatiale, temporelle), ils permettent rarement de la caractériser à l'échelle d'un individu. Plusieurs études récentes ont d'ailleurs démontré que l'hypothèse de pan-mixité est irréaliste dans la plupart des populations, ouvrant la voie à l'étude de l'impact des structures réelles de contacts sur la propagation des maladies transmissibles. La collecte de données à haute résolution sur les interactions entre individus est aujourd'hui plus facile que jamais, grâce au développement des technologies de l'information et à la miniaturisation des équipements informatiques.

Le milieu hospitalier, où des bactéries multirésistantes circulent depuis le milieu du XX<sup>ème</sup> siècle, offre un contexte de choix pour l'étude de telles données. Ces bactéries sont souvent associées à une morbidité importante dans les structures de soins. Limiter, voire endiguer la progression de ces bactéries dans la communauté, est actuellement un défi majeur de santé publique.

Dans cette thèse, nous présentons les résultats de l'analyse conjointe des contacts interindividuels et du portage individuel de *Staphylococcus aureus*. Ces données, collectées dans l'hôpital de soins de suite de Berck-sur-Mer pendant 6 mois, permettent de mettre en évidence l'intérêt de l'étude des structures de contacts pour établir des nouvelles mesures de contrôle dans la lutte contre les infections nosocomiales, et plus généralement pour apporter un nouveau substrat informatif aux modèles épidémiologiques.

### Corrélation entre portage d'une bactérie et réseaux de contacts

De plus en plus d'études utilisent des réseaux de contacts observés dans différentes situations (conférences, écoles, hôpitaux...) afin d'identifier des individus plus exposés lors de la dissémination simulée d'un pathogène. Toutefois, ces travaux font l'hypothèse que le réseau de contact tel que collecté est bien le substrat de la transmission lors d'une épidémie. En utilisant les données de l'étude I-Bird (Individual Based Investigation of Resistance Dissemination), nous avons développé une méthodologie afin de tester la corrélation entre l'incidence de *S. aureus* et le réseau des contacts entre 590 patients et

soignants. Cette approche se base sur la recherche d'infecteurs reliés, dans le réseau, aux personnes incidentes à *S. aureus* une semaine donnée. La différence de fréquence d'échantillonnage (hebdomadaire pour le portage, toutes les 30 secondes pour les contacts) et la possibilité de portage transitoire compliquent l'interprétation des chaînes de transmission supposées. Nous étudions plusieurs quantités susceptibles de mesurer la corrélation entre contacts (enregistrés par des capteurs électroniques) et colonisation bactérienne.

Nous avons tout d'abord modélisé une épidémie se propageant le long des arêtes du réseau. L'hypothèse nulle d'indépendance entre portage et contacts est simulée par une approche de Monte Carlo : les informations de colonisation sont permutées un grand nombre de fois et réparties aléatoirement entre les individus. Nous comparons la puissance de plusieurs quantités entre simulation et permutations répétées, pour finalement identifier la distribution du nombre d'intermédiaires non-colonisés entre infecteurs et infectés comme signature d'une transmission le long des arêtes du réseau. L'application du même test sur les données originales de l'étude I-Bird démontre que le réseau mesuré est bien corrélé au portage observé de *S. aureus*, confirmant pour la première fois l'hypothèse que les interactions, enregistrées par des capteurs électroniques de proximité, servent bien de support pour la transmission d'une bactérie nosocomiale.

### **Facteurs de risque de colonisation par *S. aureus* liés au réseau de contacts**

Les souches de *S. aureus* résistantes à la méticilline (SARM) sont une des causes majeures d'infections nosocomiales (infections cutanées et des voies respiratoires, notamment). En utilisant les mêmes données que précédemment, nous avons réalisé une analyse cas/témoins en comparant les propriétés structurales des réseaux de patients colonisés et non-colonisés par un SARM. Les cas et les témoins sont appariés sur la date de colonisation, afin de tenir compte de la pression de colonisation des différentes souches isolées au fil des semaines. Le lien entre colonisation et covariables explicatives est étudié par régression logistique conditionnelle, afin de tenir compte de cet appariement. Les données conjointes de portage et de contacts permettent d'identifier des facteurs de risque liés aux contacts, et soulignent le rôle crucial des interactions avec le personnel de santé.

## Thesis summary

Mathematical modeling in epidemiology has long relied on the *a priori* hypothesis of panmixia within a population. Spatial or temporal variability has generally been taken into account with structured contacts, yet rarely described at the individual scale. Recent studies have shown that panmixia is not a valid hypothesis in most populations, calling for better description of individual patterns to model the spread of infectious diseases. Purposefully, recent advances in communication technologies now allow for monitoring individual interactions at a high resolution, at relatively low costs.

A context of interest for looking at such data is the hospital, where multidrug-resistant bacteria have been circulating since the middle of the XX<sup>th</sup> century. Such bacteria are often associated with delayed healing, increased morbidity and mortality in healthcare facilities. Reducing the spread of these pathogens is and remains a major public health issue, especially now that new antibiotics are hard to identify.

In this thesis, we present the results of the joint analysis of inter-individual contacts and *S. aureus* carriage. These data were collected during a 6-month-long study in the long-term care facility in Berck-sur-Mer, France. We show that these two datasets plead for a more systematic documentation of in-hospital contact patterns, in an effort to design prevention measures targeting individuals who are at situation of increased risk to become colonized with nosocomial pathogens. In this regard, epidemiological modeling would benefit from contact data to better inform on the risk of dissemination.

### Correlation between contacts and *S. aureus* carriage

Contact networks are increasingly used to identify individuals more exposed to a theoretical pathogen during an outbreak. Such networks have been collected in various settings such as conferences, school or hospitals. However, these findings rely on the hypothesis that the observed contact network indeed captures epidemiologically relevant interactions, that is, contacts that can actually lead to transmission.

We used the data collected during the I-Bird study (Individual-Based Investigation of Resistance Dissemination) to develop a methodology to test the correlation between *S. aureus* incidence and the contact network among 590 patients and healthcare workers (HCWs). This approach is based on the network path length between incident patients and previous carriers of the same strain. To account for possibly unobserved transient carriage (due to the difference in swabbing frequency (1 week) and contact

measurement (30 s)), we allowed for « gaps » in the chain of transmission. We identified three strategies that could capture the correlation between contacts and carriage, and set out to determine whether they can be used in this respect.

At first, we used a S-I-S model to simulate the spread theoretical pathogen along network edges. The null hypothesis was achieved by a Markov Chain Monte-Carlo approach, with repeated permutations of carriage statuses. We then assessed the power of the three identified quantities. The shortest path length between an incident case and previous carriers of the same strain differs significantly, which is the signature of transmission occurring along network edges. When applied to the original I-Bird carriage data, the same behavior is observed, confirming that close proximity interaction records, measured by electronic sensors, indeed capture contacts leading to transmission of *S. aureus*.

### **Contact-oriented case-control analysis**

Methicillin-resistant *S. aureus* (MRSA) is the most common cause of nosocomial infection. We used the same data as earlier in a case-control analysis, in order to identify contact-related risk factors for MRSA colonization in patients. The analysis was nested in the longitudinal follow-up of patients to account for differential MRSA pressure depending on prevalent strains. The association between colonization and explanatory variables was investigated by conditional logistic regression. The joint analysis of contacts and carriage yielded significant associations, highlighting the role of contact networks in shaping *S. aureus* transmission course, with increased risks when contacts were HCWs.

## Productions scientifiques

### Publications

**Obadia T.**, Silhol R., Opatowski L., Temime L., Legrand J., Thiébaut A., Herrmann J.-L., Fleury É., Guillemot D., Boëlle P.-Y. (2015) Detailed Contact Data and the Dissemination of *Staphylococcus aureus* in Hospitals. PLoS Comput Biol 11(3): e1004170. doi:10.1371/journal.pcbi.1004170

**Obadia T.**, Opatowski L., Temime L., Herrmann J.-L., Fleury É., Boëlle P.-Y., Guillemot D. (2015) Inter-individual Contacts and Carriage of Methicillin-Resistant *Staphylococcus aureus*: A Nested Case-Control Study. Infection Control & Hospital Epidemiology doi :10.1017/ice.2015.89

Opatowski L., Boëlle P.-Y., Herrmann J.-L., Fleury É., Guillemot D., Labrador B., **Obadia T.**, Temime L., Thiébaut A. Estimating *S. aureus* intrinsic transmissibility in hospitals from detailed clinical and between-human contact data. En préparation

### Communications orales

**Obadia T.** (2013) « High-resolution contact network and *Staphylococcus aureus* spread in a hospital: how close-proximity interactions help identify chains of transmission ». Communication orale à EPIDEMICS 4th Conference, Amsterdam, Pays-Bas.

**Obadia T.** (2013) « Contact measurement oriented case-control analysis of acquisition risk of MRSA ». Communication orale invitée à la RICAI 2013, Paris, France.

**Obadia T.** (2014) « Les contacts d'un patient hospitalisé sont facteurs de risque de colonisation par *S. aureus* ». Communication orale au VI<sup>ème</sup> congrès de l'ADELF-EPITER, Nice, France.

**Obadia T.** (2015) « Biostatistiques, épidémiologie et réseaux de contacts ». Cours (2h) dispensé à des étudiants en Master 1 de modélisation en épidémiologie à Agrocampus-Quest, Rennes, France.

### Posters

**Obadia T.** (2013) « Detailed contact data and the dissemination of *Staphylococcus aureus* in hospitals ». Poster pour le séminaire annuel de l'École Doctorale 393, St Malo, France.

**Obadia T.** (2014) « Inter-individual contacts in transmission of methicillin-resistant *S. aureus*: a nested case-control study ». Poster pour le séminaire annuel de l'École Doctorale 393 (prix du meilleur poster), St Malo, France.

## Table des matières

Introduction .....	15
Antibiorésistance et infections nosocomiales .....	15
La modélisation, outil mathématique pour l'exploration de stratégies de contrôle .....	16
Pourquoi utiliser des modèles ? .....	17
Les modèles reposent sur des hypothèses .....	17
Données étudiées dans le cadre de cette thèse .....	18
Colonisation par <i>S. aureus</i> .....	18
Réseaux des contacts entre patients et soignants .....	18
Objectifs de cette thèse .....	18
1 Chapitre 1 : Modélisation épidémiologique et structures de contacts .....	21
1.1 Critères de transmission d'un élément pathogène .....	21
1.2 Modéliser la dynamique de transmission d'un pathogène .....	22
1.2.1 Contact menant à la transmission .....	22
1.2.2 Les modèles compartimentaux .....	22
1.2.2.1 Le modèle S-I-R .....	22
1.2.2.2 Extensions possibles du modèle S-I-R .....	23
1.2.2.2.1 Prise en compte de la démographie .....	23
1.2.2.2.2 Extensions des compartiments .....	24
1.2.2.2.3 Modèles compartimentaux structurés .....	24
1.2.2.2.4 Approche stochastique .....	25
1.2.3 Hypothèses simplificatrices du modèle S-I-R .....	25
1.2.3.1 Taux de contact dans la population .....	26
1.2.3.2 Susceptibilité .....	26
1.2.3.3 Infectiosité théorique .....	26
1.2.3.4 Transmissibilité .....	26
1.3 Pourquoi caractériser les structures de contacts de façon explicite ? .....	26
1.3.1 La limite des modèles structurés .....	26
1.3.2 Assortativité des contacts par âge : l'étude POLYMOD .....	27
1.4 Les réseaux comme structure explicite des interactions .....	31
1.4.1 Notations propres aux réseaux .....	31
1.4.1.1 Définition d'un réseau .....	31
1.4.1.1.1 Définition formelle .....	31
1.4.1.1.2 Représentation .....	32
1.4.1.2 Attributs .....	32
1.4.1.3 Mesures sur des réseaux .....	32
1.4.1.3.1 Densité .....	32
1.4.1.3.2 Degré .....	32
1.4.1.3.3 Distance minimale .....	33
1.4.1.3.4 Diamètre .....	33
1.4.1.3.5 Centralité .....	33

1.4.1.3.6	Coefficient de clustering .....	33
1.4.2	Les différents types de réseaux.....	33
1.4.2.1	Réseaux aléatoires .....	34
1.4.2.2	Réseaux petit-monde (« small-world networks ») .....	34
1.4.2.3	Réseaux invariants d'échelle (« scale-free networks »).....	35
1.5	Paramètres épidémiologiques et structures de contacts .....	36
1.5.1	L'hypothèse simple de transmissibilité densité-dépendante .....	36
1.5.2	Réseaux aléatoires et modèle S-I-S.....	37
1.5.3	Cas d'une distribution quelconque du degré .....	38
1.5.4	Prise en compte de la structure explicite des contacts.....	39
2	Chapitre 2 : Épidémiologie à l'ère du numérique.....	41
2.1	Quel type de données de contacts ?.....	41
2.1.1	Questionnaires, carnets de bord, investigateurs dédiés .....	41
2.1.2	Collecte massive et automatisée par des capteurs autonomes.....	42
2.2	Réseaux de contacts à haute résolution dans la littérature .....	44
2.2.1	Approche théorique pour caractériser des « super-transmetteurs » .....	44
2.2.2	Application pratique.....	45
2.2.2.1	Approche observationnelle.....	45
2.2.2.2	Evaluation rétroactive de mesures de contrôles informées par le réseau des contacts .....	45
2.3	Le projet I-Bird .....	46
2.3.1	Le poids des infections nosocomiales à <i>S. aureus</i> .....	46
2.3.2	Stratégies de contrôle de la transmission de bactéries nosocomiales.....	47
2.3.2.1	Contrôle direct par application des mesures d'hygiène .....	47
2.3.2.2	Contrôle indirect par l'étude des interactions interindividuelles .....	47
2.3.3	Collecte des données I-Bird et exploitations possibles .....	48
2.3.3.1	Réseau de contact à haute résolution .....	48
2.3.3.2	Suivi longitudinal du portage de <i>S. aureus</i> .....	48
2.3.4	Graphes de contacts « petit-monde » au sein de l'hôpital .....	49
2.3.4.1	Réseau complet agrégé sur l'ensemble de l'étude.....	49
2.3.4.2	Propriétés des réseaux discrétisés par jour.....	49
2.3.5	Suivi des colonisations par <i>S. aureus</i> .....	52
3	Chapitre 3 : Utilisation du réseau des contacts observés pour l'identification de chaînes de transmission de <i>S. aureus</i> en milieu hospitalier .....	53
3.1	Stratégies d'identification de chaînes de transmission.....	53
3.1.1	Traçage d'infections (« infection tracing »).....	53
3.1.2	Traçage des contacts (« contact-tracing ») .....	53
3.2	Identification de chaînes de transmission dans un réseau dynamique.....	54
3.2.1	Agrégation.....	54
3.2.2	Chemins temporellement cohérents .....	55
3.2.2.1	Existence de tels chemins .....	55
3.2.2.1.1	Cas particulier : l'information se propage d'une arête à chaque pas de temps	56



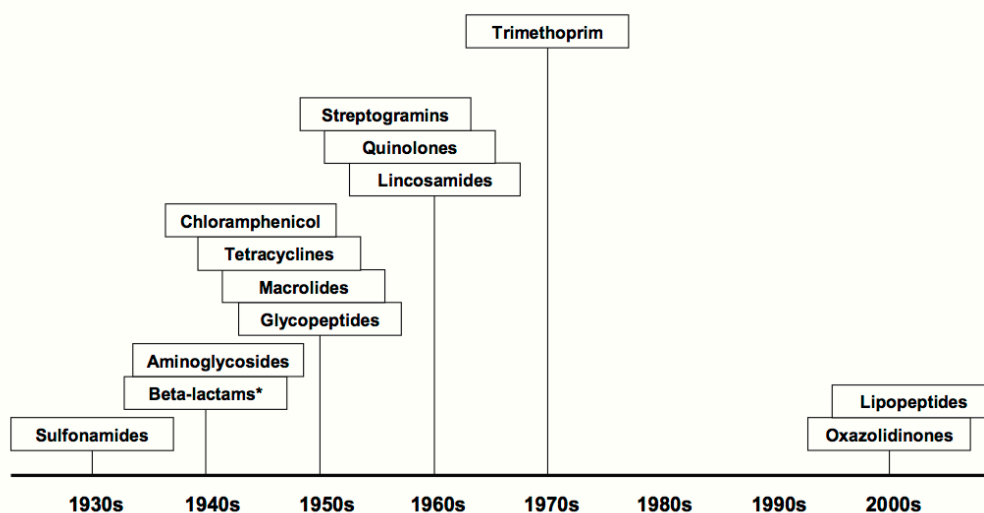
3.2.2.1.2	Cas général : l'information se propage sur un nombre variable d'arêtes à chaque pas de temps.....	56
3.2.2.2	Calcul de la longueur de ces chemins .....	57
3.3	Données de contacts détaillées et incidence de <i>S. aureus</i> en milieu hospitalier ..	58
3.3.1	Résumé de l'article en français .....	58
3.3.2	Article : Detailed Contact Data and the Dissemination of <i>Staphylococcus aureus</i> in Hospitals, Obadia T. et al., PLOS Computational Biology, 2015.....	60
3.4	Commentaires et perspectives.....	95
3.4.1	Sur les conséquences d'un réseau « petit-monde » dynamique .....	95
3.4.2	Sur l'existence du portage transitoire.....	96
3.4.3	Sur la réalité des chaînes de transmission.....	96
3.4.4	Sur la permutation du portage pour simuler l'hypothèse d'indépendance..	97
3.4.5	Sur le modèle de dissémination dans les simulations.....	97
3.4.6	Application pratique.....	98
4	Chapitre 4 : Evaluation quantitative des risques liés au réseau de contacts pour la colonisation des patients par <i>S. aureus</i> résistant à la méticilline.....	99
4.1	Facteurs de risques investigués et définition de la population .....	99
4.1.1	Types de facteurs de risques.....	99
4.1.2	Choix des cas et des témoins.....	100
4.2	Analyse cas-témoins des facteurs de risque de colonisation par SARM.....	100
4.2.1	Résumé de l'article en français .....	100
4.2.2	Article : Interindividual Contacts and Carriage of Methicillin-Resistant <i>Staphylococcus aureus</i> : A Nested Case-Control Study, Obadia T. et al., Infection Control and Hospital Epidemiology, 2015 .....	101
4.3	Commentaires et perspectives.....	111
4.3.1	Sur la représentativité du réseau mesuré .....	111
4.3.1.1	Agrégation par semaine des réseaux égo-centrés .....	111
4.3.1.2	Effet Hawthorne .....	111
4.3.2	Sur la définition des cas.....	111
4.3.3	Sur les facteurs de risque identifiés.....	112
4.3.3.1	Peu de significativité pour les colonisations par <i>S. aureus</i> sensible à la méticilline.....	112
4.3.3.2	Facteurs de risque « réseau » .....	113
4.3.3.3	Facteurs de risque « réseau et microbiologie » .....	113
5	Chapitre 5 : Dynamiques de formation et de dissolution de liens et considérations épidémiologiques dans un réseau à topologie variable.....	115
5.1	Données manquantes et méthodologie d'échantillonnage du réseau .....	115
5.1.1	Méthode d'échantillonnage des contacts.....	116
5.1.2	Notations .....	116
5.1.3	Inférence pour un réseau résultant d'un modèle paramétrique.....	116
5.2	Pistes pour l'imputation des données manquantes.....	117
5.2.1	Imputation par la description du voisinage (approche empirique) .....	117
5.2.2	Imputation par modélisation de la formation des liens.....	119

5.2.2.1 Les ERGM (Exponential Random Graph Models) .....	119
5.2.2.1.1 Écriture du modèle paramétrique pour $Y$ .....	119
5.2.2.1.2 Description des statistiques contenues dans $g(y)$ .....	119
5.2.2.2 Imputation des liens manquants par des approches de calcul intensif .....	120
5.2.3 Exemple .....	120
5.2.4 Separable Temporal ERGM : extension aux réseaux dynamiques .....	123
5.3 Perspectives épidémiologique de prédiction du risque épidémique .....	123
5.3.1 Extension du modèle S-I-S sur un substrat dynamique .....	123
5.3.2 Perspectives d'application aux données I-Bird .....	124
5.3.2.1 Evaluation de seuils épidémiques .....	124
5.3.2.1.1 Seuil épidémique selon les contacts .....	124
5.3.2.1.2 Variations dans le temps .....	125
5.3.2.1.3 Variations dans les différents service de l'hôpital .....	125
5.3.2.2 Vers des simulations d'interventions sur la structure du réseau .....	126
Conclusion .....	127
Sur l'hétérogénéité des contacts .....	127
Sur la faisabilité de la collecte fine de réseaux de contacts .....	127
Sur la caractérisation des contacts « à risque » .....	128
Perspectives pour l'élaboration de mesures de contrôle informées par un réseau de contacts .....	129
Bibliographie .....	131
Annexe 1 : Résultats de l'analyse cas-témoins pour la colonisation des patients par SARM en utilisant des données lissées .....	145
Annexe 2 : Résultats de l'analyse cas-témoins pour la colonisation des patients par SASM .....	149
Annexe 3 : Dictionnaire des covariables utilisées aux Annexes 1 et 2 .....	153

## Introduction

### Antibiorésistance et infections nosocomiales

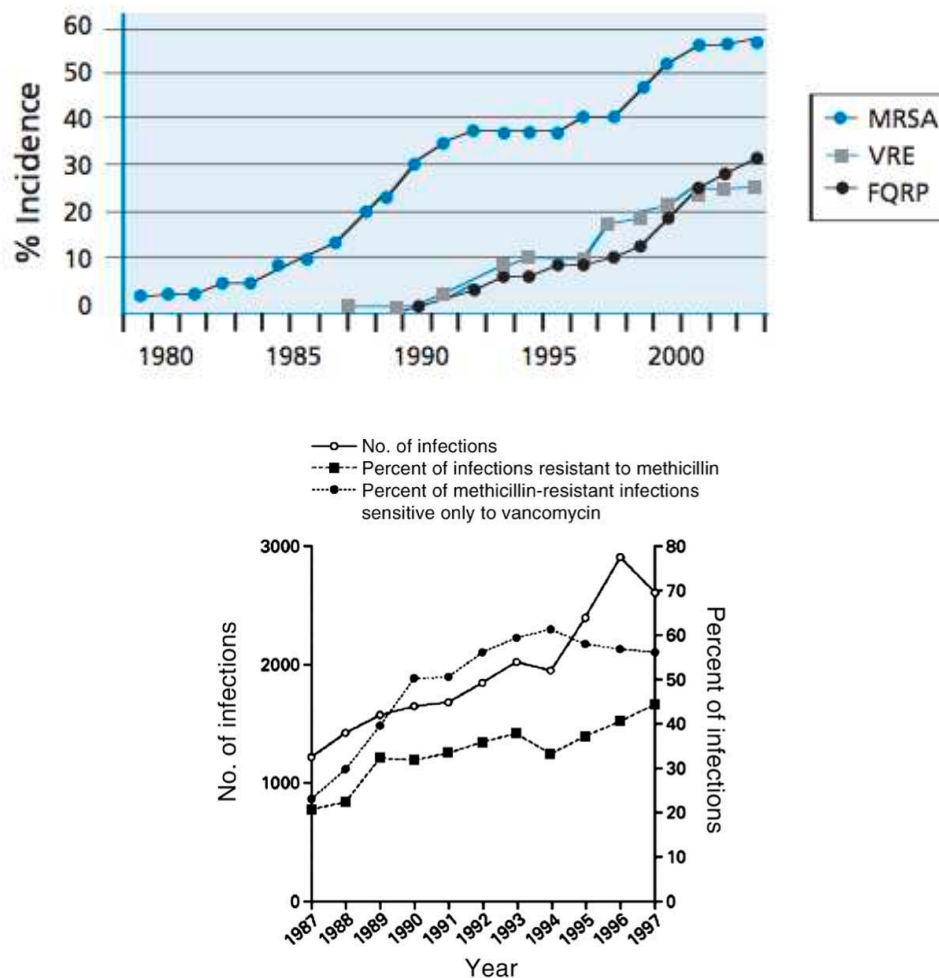
Au cours du XX<sup>ème</sup> siècle, de nombreuses classes d'antibiotiques ont été découvertes à un rythme soutenu. Le repère historique le plus marquant est sans doute la découverte de la pénicilline en 1928.[53] Cependant, depuis la fin des années 1970, peu de nouveaux antibiotiques ont été identifiés[49] (voir Figure 1) et certaines bactéries ont développé des mécanismes de résistance aux antibiotiques utilisés traditionnellement en première intention. A titre d'exemple, les premières résistances à la pénicilline sont observées dès le milieu des années 1940.[89]



**Figure 1. Découverte de nouvelles classes d'antibiotiques.** Figure reproduite de European Center for Disease Prevention and Control. (2009)

L'absence de nouveaux antibiotiques est d'autant plus préoccupante que la dissémination géographique des souches résistantes est assez rapide.[25, 135] Pour certaines bactéries, la résistance aux antibiotiques devient presque une norme : parmi elles, *Staphylococcus aureus* occupe une place de choix (voir Figure 2), responsable d'un grand nombre d'infections nosocomiales diverses, et ayant rapidement acquis une résistance à la méticilline.[98, 110] L'incidence croissante de ces souches multirésistantes devient une problématique majeure de santé publique : dans certains pays comme aux Etats-Unis, certaines souches résistantes se trouvent dans la communauté. En Europe, les souches de *S. aureus* résistantes à la méticilline sont encore majoritairement acquises à l'hôpital. L'enjeu est double : il est nécessaire de trouver des

moyens efficaces pour réduire le poids des infections nosocomiales, ce qui permet également de limiter la dissémination des souches responsables dans la communauté.



**Figure 2. Évolution de l'incidence de souches multirésistantes. (Haut)** Incidence relative de *S. aureus* résistant à la méticilline (MRSA), des entérocoques résistants à la vancomycine (VRE) et de *Pseudomonas aeruginosa* résistant aux fluoroquinolones (FQRP). Figure reproduite de Infectious Diseases Society of America. (2004) **(Bas)** Incidence annuelle de *S. aureus* résistant à la méticilline. Figure reproduite de Lowy et al. (2003)

*S. aureus* est une bactérie commensale : les infections nosocomiales à *S. aureus* sont précédées par un épisode de colonisation. Réduire la colonisation par ces bactéries multirésistantes est la voie naturelle vers la réduction de l'incidence des infections nosocomiales qu'elles peuvent entraîner.

### La modélisation, outil mathématique pour l'exploration de stratégies de contrôle

L'émergence de mécanismes de résistance est liée au processus d'évolution et de sélection naturelle. Un premier axe d'étude est de développer des outils pour détecter au plus tôt l'émergence de ces mécanismes. Nous nous intéresserons dans cette thèse à un autre volet dans la lutte contre la dissémination de l'antibiorésistance, celui de la

planification de mécanismes de contrôle une fois qu'une souche résistante est détectée. Lorsqu'un patient admis à l'hôpital est colonisé par une bactérie multirésistante, des procédures spécifiques sont appliquées, afin de l'isoler au mieux et de limiter les risques de dissémination à d'autres individus. Ces mesures se traduisent par des recommandations d'hygiène accrues pour le personnel soignant (port de surblouses, gants, limitation de déplacements du patient, utilisation de solutions hydro-alcooliques...), et ont fait l'objet d'un réel défi de santé publique.[118]

### **Pourquoi utiliser des modèles ?**

Evidemment, il n'est pas possible de tester à l'hôpital l'efficacité de toutes les actions possibles. La modélisation mathématique devient alors un outil indispensable, permettant d'évaluer l'impact de ces mesures dans un cadre théorique, parfois paramétré par des données observées. En épidémiologie, de nombreux modèles mathématiques sont utilisés afin de simuler de façon déterministe ou stochastique l'évolution d'un agent infectieux dans une population. Historiquement, le modèle le plus célèbre est probablement celui développé par Kermack & McKendrick,[85] dit « modèle S-I-R » (Susceptible-Infectieux-Retiré) : la population d'intérêt est découpée selon des classes reflétant l'évolution naturelle de la maladie, et les individus sont infectés par des « contacts » avec des individus infectieux.

La comparaison entre des modèles et des données permet de comparer leur performance. Lorsqu'un modèle reproduit fidèlement l'évolution d'une maladie, il peut alors être altéré, afin de prendre en compte des mesures de contrôles. Des simulations permettent alors d'évaluer *a priori* l'impact de ces mesures, et éventuellement de comparer leur performance. Une bonne évaluation nécessite bien entendu un modèle capable de représenter fidèlement les mécanismes de transmission de la maladie d'intérêt : les paramètres d'entrée du modèle sont une donnée cruciale.

### **Les modèles reposent sur des hypothèses**

Les modèles mathématiques font souvent des postulats afin de simplifier leur écriture. Dans le cas du modèle S-I-R « historique », il s'agit de l'hypothèse d'homogénéité de la population : à l'intérieur de chaque compartiment, les individus sont tous égaux face au pathogène, et dans leurs comportements. Ils sont supposés mélangés de façon homogène. L'intuition veut que cette hypothèse soit peu réaliste. Dans cette thèse, nous allons nous intéresser à une source d'hétérogénéité qui suscite des travaux croissants en

épidémiologie : l'hétérogénéité des contacts entre individus. Il est raisonnable d'envisager que tous les individus d'une population n'ont pas la même probabilité d'entrer en contact (éloignement géographique, affinité par tranche d'âge et/ou par sexe...), et ces différences de comportements peuvent affecter le déroulement d'une épidémie dans la population.[83] La caractérisation des contacts, au sens épidémiologique du terme, est souvent difficile, et des mesures de proxy sont utilisées.[42]

### **Données étudiées dans le cadre de cette thèse**

Dans ce travail, nous utiliserons les données de l'étude I-Bird (Individual-Based Investigation of Resistance Dissemination). Ces données ont été collectées dans l'hôpital de soins de suites et de réadaptation de Berck-sur-Mer en 2009.

### **Colonisation par *S. aureus***

Comme évoqué précédemment, *S. aureus* est responsable d'un nombre croissant d'infections nosocomiales, et il en existe de nombreux variants résistants. Nous disposons ici de données hebdomadaires de colonisation de tous les patients et soignants de l'hôpital dans lequel s'est déroulée l'étude.

### **Réseaux des contacts entre patients et soignants**

La caractérisation des contacts à risque de transmission d'une bactérie est délicate. *S. aureus* se transmettant principalement lors des contacts physiques, nous avons choisi d'enregistrer le réseau des interactions entre tous les participants à l'étude : il s'agit d'un graphe de contacts dynamique, qui reflète la proximité géographique des individus, ainsi que la durée de cette proximité.

### **Objectifs de cette thèse**

Au fil de cette introduction, nous voyons apparaître plusieurs points essentiels interconnectés : la dissémination de l'antibiorésistance est une problématique de santé publique où la modélisation mathématique peut permettre d'identifier des stratégies de contrôle efficaces. La mise au point d'un modèle capable de représenter fidèlement les processus menant à la transmission des souches qui nous intéressent nécessite au préalable d'en comprendre les mécanismes de transmission. Dans le cas de *S. aureus*, l'hypothèse que les contacts entre individus sont en lien avec l'incidence de la bactérie est plausible *a priori*, mais n'a pas été vérifiée, faute de données suffisantes.

Dans le Chapitre 1, nous passerons en revue différents modèles épidémiologiques, et évoquerons les hypothèses qu'ils imposent. L'étude de ces modèles permettra d'identifier des quantités qui reflètent la probabilité de survie ou d'extinction d'un pathogène, qui est un critère utile dans l'évaluation de stratégies de contrôle. Nous évoquerons également plusieurs méthodes qui ont déjà été utilisées afin de s'affranchir (au moins en partie) de l'hypothèse d'homogénéité interindividuelle.

Au Chapitre 2, nous discuterons des possibilités offertes afin de mesurer l'hétérogénéité des contacts dans une population. En effet, les progrès majeurs réalisés dans le domaine des technologies de l'information depuis les années 1990 permettent de mesurer plus finement et plus efficacement les comportements individuels. Ce « big data » épidémiologique pave la voie vers une réflexion presque individualisée en termes de précautions, et dans ce contexte nous présenterons en détails l'étude I-Bird et les données qui y ont été collectées.

Au Chapitre 3, nous utiliserons ces données afin de légitimer une hypothèse utilisée de façon empirique dans « l'épidémiologie numérique ». En effet, les réseaux de contacts sont de plus en plus utilisés afin de représenter l'hétérogénéité du substrat servant à la transmission d'une maladie. Toutefois, il n'a jamais été possible de tester la corrélation entre l'incidence de cette maladie et les contacts utilisés. Nous développons une méthodologie statistique, basée sur des permutations répétées, afin de tester cette corrélation. Ce chapitre sera aussi l'occasion d'aborder la problématique des réseaux dynamiques, dans lesquels la corrélation temporelle des liens est à prendre en compte. Nous verrons que le réseau mesuré dans I-Bird est bien en accord avec l'hypothèse de transmission de *S. aureus* au cours des contacts entre les participants.

Dans la continuité de cette démarche, le Chapitre 4 utilisera ces données de contacts afin de déterminer des facteurs de risque pour la colonisation de patients par *S. aureus*. Si la méthodologie est plus traditionnelle, nous l'appliquons à des données de contacts, afin de mieux caractériser les interactions entre individus qui sont susceptibles de mener à la transmission entre un individu colonisé et un individu non colonisé.

Enfin, nous proposerons dans le Chapitre 5 des pistes de réflexion pour répondre aux problématiques identifiées au cours de ce travail. La thématique des données manquantes, récurrente en statistiques, sera l'occasion d'introduire des méthodes de modélisation de réseaux sociaux. Nous explorerons plus en détail les questions soulevées par l'exploitation de réseaux dynamiques, en proposant des méthodes

mathématiques qui pourront permettre de tester l'efficacité de diverses mesures de prévention, informées par le réseau des contacts dans la population étudiée.



# **1 Chapitre 1 : Modélisation épidémiologique et structures de contacts**

Dans ce chapitre, nous nous intéressons d'abord aux principes fondamentaux de la modélisation épidémiologique. Nous définissons dans un premier temps la méthodologie usuelle pour la modélisation de la transmission d'une infection, en séparant une population en plusieurs catégories d'individus. Nous présentons le modèle historique Susceptibles-Infectieux-Retirés (S-I-R), et abordons plusieurs extensions possibles. Par la suite, nous discutons le réalisme des hypothèses faites par ces modèles sur la structure des contacts dans les populations. Enfin, nous passons en revue les travaux déjà menés pour évaluer l'impact de la structure des contacts sur un modèle de transmission.

## **1.1 Critères de transmission d'un élément pathogène**

Dans le contexte des maladies transmissibles, les modèles de transmission cherchent à reproduire les épisodes de survenue de la maladie chez les personnes la contractant après contamination (incidence). Il est donc nécessaire de distinguer l'agent infectieux (pathogène) de l'observation directe de la maladie. Le potentiel d'infectiosité d'un individu malade dépend d'une combinaison de facteurs intrinsèques (quantité de pathogène excrété, statut immunitaire...) et extrinsèques (capacité du pathogène à survivre à l'extérieur de l'hôte, voies de transmission du pathogène, nombre et caractéristiques des interactions de la personne malade avec d'autres personnes susceptibles...). Ces facteurs peuvent être corrélés.[65] Pour rendre possible un événement de transmission entre deux individus A et B, il faut donc réunir 4 critères :

- L'individu A est infectieux et excrète du pathogène selon une certaine dynamique,
- Le pathogène excrété survit en dehors de son hôte un temps suffisant,
- L'individu B entre en contact directement (contact physique avec A) ou indirectement (environnement contaminé) avec le pathogène,
- Le pathogène survit et prolifère chez l'individu B.

Nous nous intéresserons ici au cas général d'une transmission possiblement « silencieuse », c'est à dire n'entraînant pas forcément la survenue d'une maladie : c'est notamment le cas du portage d'une bactérie nosocomiale, où l'incidence sera caractérisée par l'acquisition de la bactérie.

## **1.2 Modéliser la dynamique de transmission d'un pathogène**

Plusieurs voies sont à envisager dans le cas général de transmission. Un individu infectieux peut contaminer son environnement, soit par contact physique soit par l'excrétion de pathogène sous la forme de particules aéroportées. Ces vecteurs passifs sont aussi appelés « fomites ». Nous nous intéresserons au cas plus général d'une transmission par contact.

### **1.2.1 Contact menant à la transmission**

La notion de "contact" est intimement liée aux modalités de dissémination du pathogène étudié. Par exemple, dans le cas des infections sexuellement transmissibles, ces contacts sont sexuels ou sanguins.[124, 125] Pour des maladies transmissibles, les contacts font appel à la notion de proximité dans l'espace et le temps, entre un individu infecté et un individu susceptible. En 1906, Hamer propose la loi d'action de masse, dont l'utilisation classique établit le postulat que l'incidence d'une maladie infectieuse résulte des contacts entre les individus susceptibles de la contracter et ceux pouvant la transmettre. Souvent, les contacts amenant réellement à transmission ne sont pas mesurables, et une définition alternative est proposée comme proxy.

### **1.2.2 Les modèles compartimentaux**

Si les contacts sont vecteurs de transmission, ils ne capturent pas l'état dans lequel un individu se trouve à un instant donné. Élaborés au début du XX<sup>ème</sup> siècle et découlant de la loi d'action de masse, les premiers modèles compartimentaux répartissent les individus d'une population dans des compartiments traduisant leur état vis-à-vis du pathogène considéré. Les compartiments ainsi établis permettent de décrire, à chaque instant, la répartition de la population selon le critère infectieux d'intérêt.

#### **1.2.2.1 Le modèle S-I-R**

Introduit en 1916 pour la malaria[127] puis généralisé quelques années plus tard[85–87], ce modèle permet de classer les individus selon qu'ils sont Susceptibles, Infectieux, ou Retirés de la dynamique épidémique selon le schéma proposé en Figure 3.



**Figure 3. Modèle compartimental Susceptible (S), Infectieux (I), Retiré (R).** Les paramètres  $\beta$  et  $\gamma$  correspondent aux probabilités de transition d'un compartiment au suivant.

La transition du compartiment S au compartiment I fait intervenir un paramètre de transmissibilité ( $\beta$ , qui capture le produit de l'infectiosité et du taux de contact entre individus), et les individus sortent du compartiment I à un taux  $\gamma$ , correspondant à l'inverse de la période infectieuse.

L'hypothèse la plus répandue est que la force d'infection dépend non pas du nombre absolu d'individus infectieux mais de leur proportion dans la population.[5]

Dès lors, le modèle est gouverné par le système d'équations différentielles :

$$\begin{cases} \frac{dS}{dt} = -\frac{\beta SI}{n} \\ \frac{dI}{dt} = \frac{\beta SI}{n} - \gamma I \\ \frac{dR}{dt} = \gamma I \end{cases}$$

Avec cette hypothèse de transmission fréquence-dépendante, on remarque que la dynamique épidémique dépend du ratio  $R_0 = \frac{\beta}{\gamma}$ , aussi appelé taux de reproduction de base. C'est le théorème du seuil: une épidémie ne peut démarrer que si la fraction initiale d'individus susceptibles est plus grande que  $\frac{1}{R_0}$ . [5, 84] Dans une population presque entièrement susceptible à l'état initial, ce seuil revient à comparer  $R_0$  à la valeur 1 :  $R_0$  représente le nombre moyen d'individus que va infecter un individu infectieux au cours de sa période infectieuse, dans une population entièrement susceptible.

### **1.2.2.2 Extensions possibles du modèle S-I-R**

Si l'expression analytique précédente de  $R_0$  n'est valable que dans le cadre du modèle S-I-R, il est possible d'adapter les compartiments et les flux du modèle pour représenter plus fidèlement la dynamique d'un pathogène.

#### **1.2.2.2.1 Prise en compte de la démographie**

Les processus de naissance et de mort peuvent être pris en compte, en générant un afflux de nouveaux individus dans la classe des susceptibles (naissances), et un retrait d'individus dans les trois compartiments (morts).

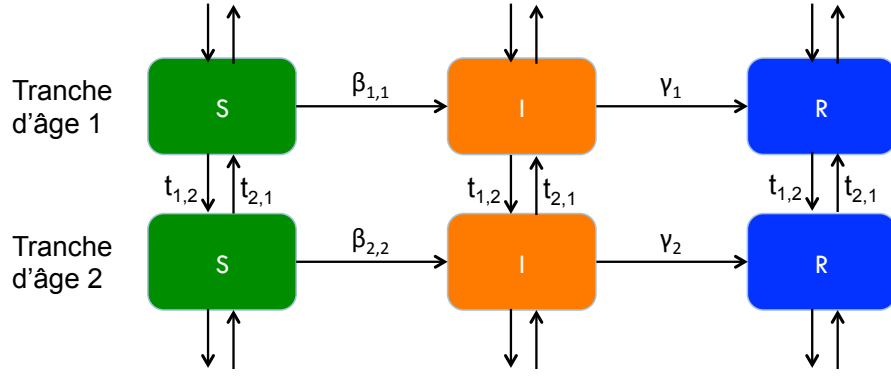
#### 1.2.2.2.2 Extensions des compartiments

L'histoire naturelle de certaines maladies n'est pas compatible avec les seules classes S-I-R. La prise en compte d'une période de latence, où un individu a déjà été infecté mais n'est pas encore infectieux, est théorisée dans le modèle S-E-I-R. Ce modèle est adapté aux maladies transmissibles pour lesquelles les symptômes ne sont pas apparents immédiatement, comme la grippe.[5, 29, 52] Ce compartiment de latence produit une dynamique plus lente, mais plus réaliste. Il est également possible de modéliser une immunité temporaire par une classe ne pouvant pas être infectée mais où l'immunité ne dure qu'une certaine période (modèles M-S-(E)-I-R où M représente l'immunité maternelle, utilisés notamment pour la rougeole et la varicelle). L'absence totale d'immunité est elle aussi envisageable, comme dans le modèle S-I-S, où les individus sortant du compartiment infectieux redeviennent susceptibles.

Pour les maladies où la période infectieuse dure longtemps à l'échelle de la vie humaine, les migrations ainsi que les processus démographiques de naissance et de mort peuvent également être inclus.[5, 84] Enfin, l'extension des compartiments peut également représenter l'implémentation de mesures de contrôles (vaccination, quarantaine,...).[66]

#### 1.2.2.2.3 Modèles compartimentaux structurés

Lorsque la population est hétérogène vis-à-vis d'un critère pouvant influencer sur le cours de la maladie, il est possible de la classer non-seulement selon l'état infectieux des individus, mais également selon le(s) critère(s) d'hétérogénéité identifiés. Les transitions entre les sous-populations caractérisant l'hétérogénéité sont alors à rajouter aux processus de transition entre les compartiments épidémiologiques. Dans le cas d'un processus épidémiologique de type S-I-R, et dans une population découpée en  $k$  groupes d'hétérogénéité (par exemple, des tranches d'âges), le modèle général est celui présenté Figure 4 (pour  $k = 2$ ).



**Figure 4. Modèle compartimental S-I-R structuré en classes d'âges.** Deux classes sont représentées, avec transition entre elles.

Ce modèle est gouverné par le système d'équations différentielles :

$$\forall i \in \{1, \dots, k\} \begin{cases} \frac{dS_i}{dt} = - \sum_{j=1}^k \frac{\beta_{ji} S_i I_j}{N_i} - \sum_{j \neq i} t_{ij} S_i + \sum_{j \neq i} t_{ji} S_j \\ \frac{dI_i}{dt} = \sum_{j=1}^k \frac{\beta_{ji} S_i I_j}{N_i} - \gamma_i I_i - \sum_{j \neq i} t_{ij} I_i + \sum_{j \neq i} t_{ji} I_j \\ \frac{dR_i}{dt} = \gamma_i I_i - \sum_{j \neq i} t_{ij} R_i + \sum_{j \neq i} t_{ji} R_j \end{cases}$$

Dans ce modèle,  $\beta_{ji}$  représente le taux de contact efficace entre les individus de la classe  $i$  et ceux de la classe  $j$ ,  $\gamma_i$  est l'inverse de la période infectieuse dans la classe  $i$ , et  $t_{ij}$  le taux de transfert de la classe  $i$  vers la classe  $j$ .

#### 1.2.2.2.4 Approche stochastique

Les modèles présentés jusque-là existent, en plus de leur approche déterministe, dans une version stochastique : le vecteur décrivant l'état épidémiologique à un instant  $t$  est Markovien, et l'incidence cumulée ainsi que le nombre d'individus retirés sont des processus de comptage. On peut formuler analytiquement la probabilité d'incrémenter l'une ou l'autre de ces variables de comptage en fonction des paramètres du modèle.

### 1.2.3 Hypothèses simplificatrices du modèle S-I-R

Par construction, les modèles compartimentaux offrent un cadre modulable pour la description de la dynamique épidémiologique dans une population pouvant être hétérogène. Toutefois, à l'intérieur d'un compartiment, l'hypothèse d'un mélange

homogène des individus est faite. Nous résumons ici les hypothèses communes faite dans ce modèle « historique ».

#### **1.2.3.1 Taux de contact dans la population**

Toutes les paires d'individus ont la même probabilité de contact, indépendamment du statut infectieux et donc de la dynamique d'excrétion (ou des symptômes). Le taux de contact ne dépend également pas du temps.

#### **1.2.3.2 Susceptibilité**

A l'intérieur d'un compartiment, tous les individus partagent la même susceptibilité vis-à-vis du pathogène. Pour chaque individu, elle vaut 1 si il est dans la catégorie S, 0 sinon.

#### **1.2.3.3 Infectiosité théorique**

De la même façon, l'infectiosité théorique est la même pour tous les individus d'un même compartiment : elle est non-nulle uniquement dans le compartiment des infectieux.

#### **1.2.3.4 Transmissibilité**

Le paramètre  $\beta$  du modèle S-I-R peut être défini comme le produit du taux de contact dans la population et de l'infectiosité théorique : c'est en général une constante qui ne tient pas compte des variabilités individuelles.

Ces hypothèses s'appliquent quel que soit le modèle considéré. Les modèles structurés permettent une plus grande variabilité des paramètres, mais dans chaque sous-population (déterminée par un critère d'hétérogénéité dans la population totale) les individus restent virtuellement identiques à l'intérieur de chaque compartiment.

### **1.3 Pourquoi caractériser les structures de contacts de façon explicite ?**

#### **1.3.1 La limite des modèles structurés**

Dans les modèles compartimentaux structurés, l'assortativité des contacts des différentes classes d'hétérogénéité est résumée sous forme d'une matrice de dimension  $(k \times k)$ , où  $k$  représente le nombre de groupes d'hétérogénéité, et où l'élément  $\beta_{j,i}$  correspond à la probabilité de contact d'un individu de la classe  $j$  avec un individu de la classe  $i$ . Ce type de matrice est communément appelé « Who Acquires Infection From Whom matrix » et permet de résumer partiellement l'hétérogénéité des individus : on

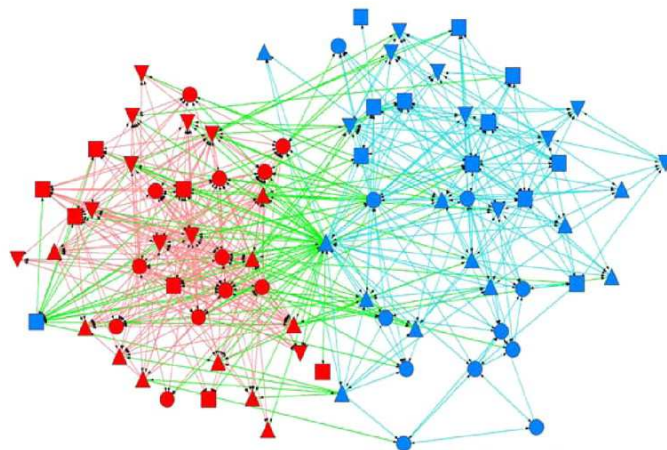
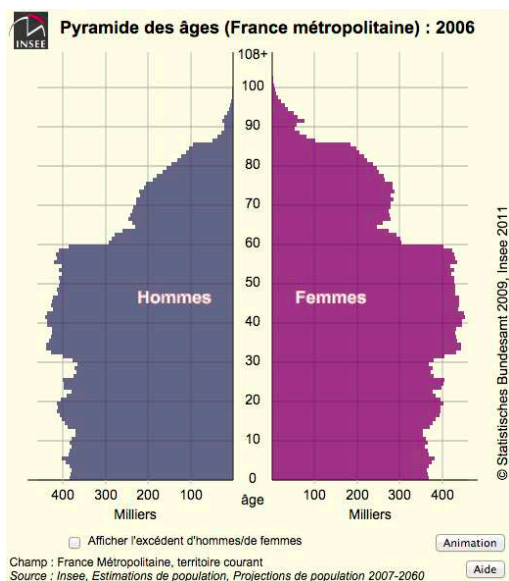
peut calculer les forces d'infections pour chaque classe d'hétérogénéité en intégrant le produit  $\beta_{i,j} * I_j(t)$  sur toutes les classes disponibles. Ce type d'approche a été utilisé pour étudier la propagation de maladies, notamment la grippe,[109] et pour tester des stratégies de vaccination ciblant les individus à risque élevé de complications (modèle SEIR stratifié par âge, utilisant des données de contacts collectées en 1986.[40, 148] Le taux de contact pourrait également être variable dans le temps.

Toutefois, il s'agit ici de paramétrer le modèle *a priori*, et les estimations des paramètres de reproduction sont largement affectées par la structure de ces matrices.[51, 145] Les prédictions de ces modèles sont dépendantes des données (souvent démographiques) qui servent à établir les critères d'hétérogénéité. De plus, ce type de modèle ne permet pas de s'affranchir des différences individuelles, notamment comportementales, dans l'établissement d'un réseau de contacts pouvant conduire à la transmission du pathogène.

Des modèles de métapopulation, ou d'autres individus-centrés, ont été utilisés pour moduler le taux de contact entre individus (ou entre métapopulations), en faisant par exemple l'hypothèse que la probabilité de contact entre deux entités décroît lorsque la distance qui les sépare augmente.[52] Les modèles gravitaires tiennent compte de la distance entre les individus et également de la distance entre leur métapopulation de rattachement.[46, 156] Enfin, d'autres travaux ont montré l'importance des réseaux de transports (aériens, nationaux, voire urbains) sur la diffusion d'une épidémie, et leur intérêt dans la mise en place de mesures de contrôle.[6, 27, 128]

### **1.3.2 Assortativité des contacts par âge : l'étude POLYMOD**

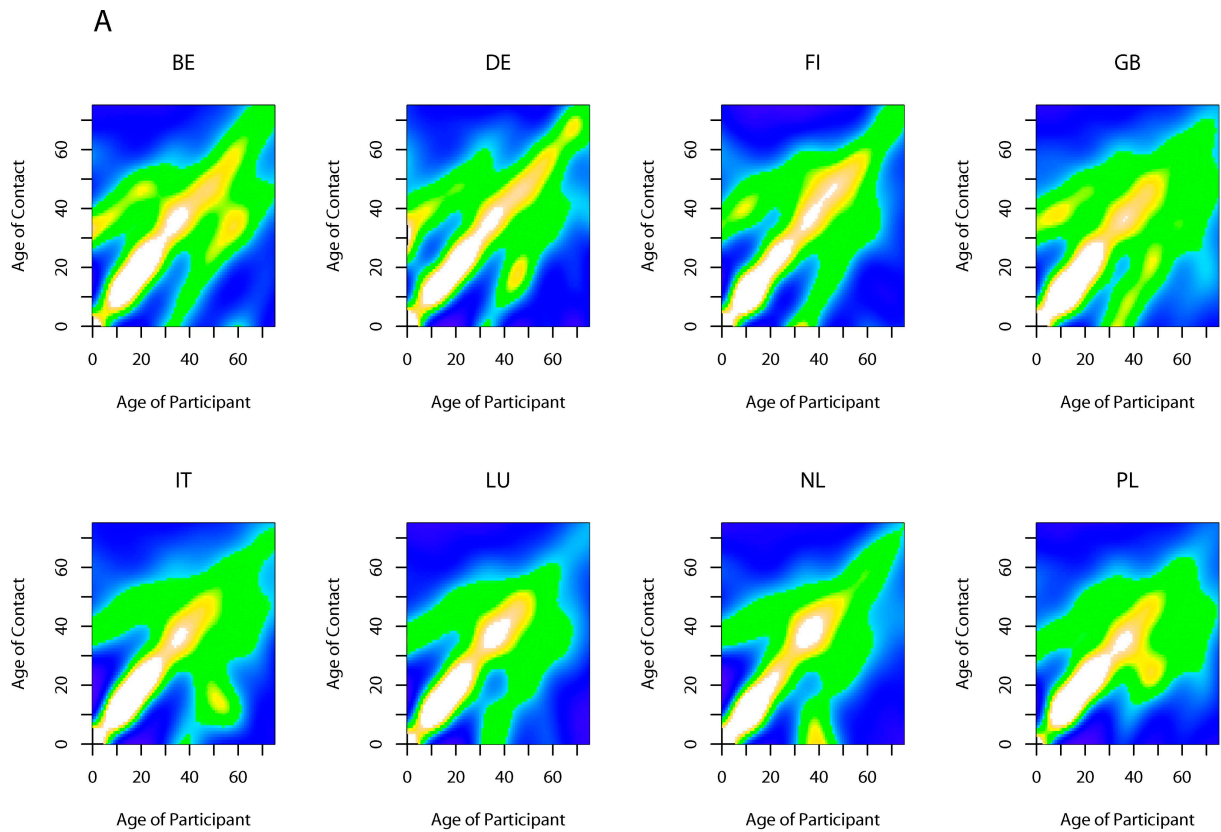
L'hypothèse d'homogénéité de la population faite dans le modèle S-I-R semble peu plausible *a priori*. Il est assez intuitif d'imaginer que toutes les paires d'individus d'une population n'ont pas toutes la même probabilité de se rencontrer. L'exemple le plus parlant est la stratification par âge de la population, et sa répartition au sein des foyers : en observant la Figure 5, il paraît peu probable que les individus de classes d'âge différentes aient tous la même probabilité de contact. Un autre exemple d'homophilie plausible peut être envisagé sur un critère de sexe, au moins dans certaines classes d'âge, comme observé en milieu scolaire.[21]



**Figure 5. Deux sources d'hétérogénéité des contacts. (Gauche)** Pyramide des âges de la population française pour l'année 2006 (source : INSEE). **(Droite)** Réseau des camarades de classe d'enfants scolarisés aux Etats-Unis (4th grande, équivalent au CM1). Chaque forme indique une classe différente, les couleurs donnent le sexe des enfants (bleu : garçon ; rouge : fille). L'algorithme de placement des nœuds ne tient pas compte du sexe pour la réalisation du graphe. Figure adaptée de Cauchemez et al. (2009)

La première étude à grande échelle des interactions sociales entre individus a été menée en 2008.[108] Elle documente de façon très précise la répartition des contacts, par âge, dans des échantillons représentatifs de la population de 8 pays européens (Belgique, Allemagne, Finlande, Grande Bretagne, Italie, Luxembourg, Pays Bas, Pologne). Les résultats de cette étude montrent une grande similarité dans la répartition des contacts selon l'emplacement où ils ont lieu, ainsi qu'une similarité dans les durées, fréquences et nombres des contacts. Le résultat le plus marquant est la claire assortativité par tranche d'âges (propension à avoir des contacts préférentiellement avec certains individus plutôt que d'autres), comme visible dans la Figure 6.





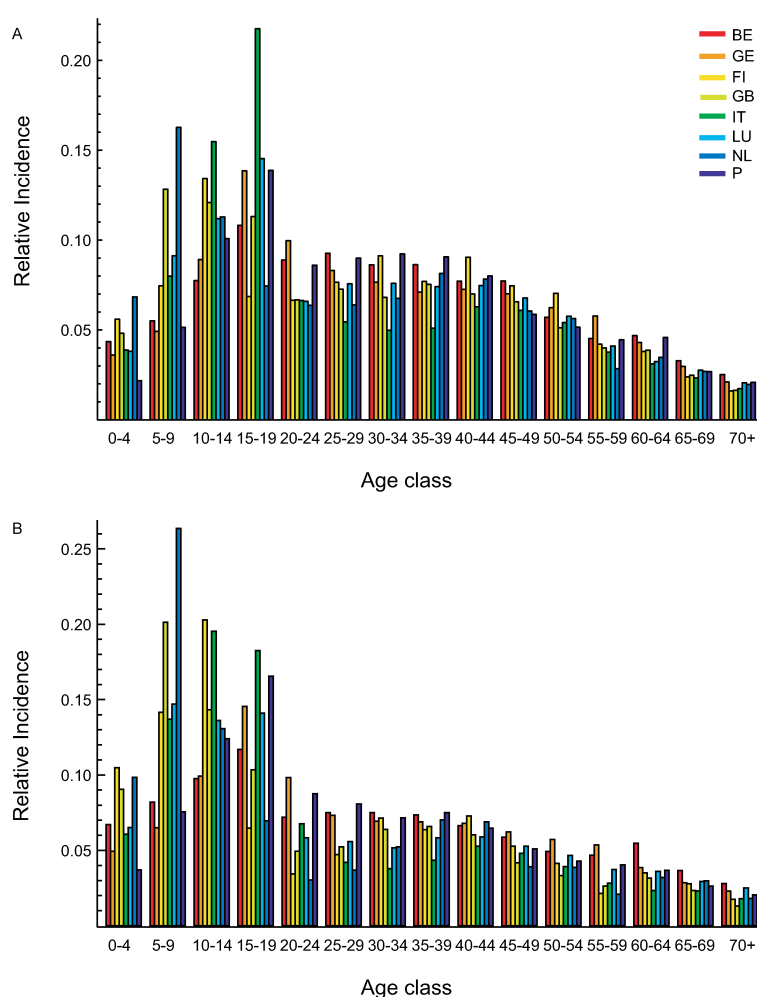
**Figure 6. Matrices de contacts lissées de l'étude POLYMOD pour l'ensemble des interactions recueillies.** Chaque encart représente les données d'un pays. Les taux les plus élevés sont en blanc et les plus faibles en bleu. Figure adaptée de Mossong et al. (2008)

Ces matrices de contacts illustrent la tendance des individus à établir des contacts principalement avec d'autres personnes ayant approximativement le même âge qu'eux (première bissectrice), ainsi qu'avec des gens d'une génération plus âgée (diagonale supérieure, génération des « grands-parents ») et d'une génération plus jeune (diagonale inférieure, génération des « enfants »). Les contacts fréquents apparaissent en blanc, ceux moyennement fréquents en vert, et ceux rares en bleu. Par construction, ces matrices ne sont pas symétriques, car elles utilisent des données égocentriques et indiquent donc, pour un individu d'âge  $i$  (abscisse), la fréquence de ses contacts avec toutes les classes d'âges possibles (ordonnée).

Les auteurs utilisent également ces données afin d'informer un modèle de transmission sur la structure réelle des contacts dans la population : ils introduisent une « Next Generation Matrix »  $K$  (inspirée de Diekmann & Hesterbeek[36]) qui représente la propension des individus à établir des contacts entre eux, pour chaque couple de classe d'âge. En notant  $X$  le vecteur contenant le nombre de cas incidents, l'incidence moyenne dans chaque strate, après  $i$  générations, est alors :

$$X(i) = K^i * X(0)$$

En la stratifiant sur 15 classes d'âges, les auteurs identifient les enfants (classe 5-9 ans) comme la catégorie avec le pic d'incidence le plus élevé. Une cause plausible est leur grand nombre de contacts, principalement en milieu scolaire, avec d'autres enfants du même âge. Ce résultat est également en accord avec une autre étude sur la propagation de la souche grippale pandémique H1N1 dans une école aux Etats-Unis, où les liens reliaient principalement des enfants de même sexe, qui transmettaient donc préférentiellement à d'autres individus du même sexe (voir Figure 5).[21] De plus, une distinction apparaît entre les contacts physiques et les interactions de proximité sans contact physique (Figure 7).



**Figure 7. Incidence relative d'une maladie transmissible dans une population totalement susceptible à l'état initial, et où l'assortativité des contacts par âge suit celle observée dans l'étude POLYMOD.** Figure reproduite de Mossong et al. (2008)

Ces données viennent de nouveau appuyer le besoin d'informer les modèles épidémiologiques sur la structure des contacts des populations: les hypothèses

simplificatrices d'homogénéité (contacts, transmissibilité) ne sont pas réalistes dans une population manifestement hétérogène.

#### **1.4 Les réseaux comme structure explicite des interactions**

Dans cette partie, nous formalisons l'utilisation des réseaux afin de décrire les contacts entre les individus.

##### **1.4.1 Notations propres aux réseaux**

Cette section s'inspire de l'ouvrage de Watts & Strogatz[152] et de l'article de Albert & Barabási[1]. Les réseaux permettent de représenter visuellement des données relationnelles, en représentant des entités (nœuds) reliées par des arêtes lorsqu'elles partagent une relation d'intérêt. Ils offrent un cadre applicable à des données variées, allant des interactions sociales aux réseaux de régulations génétiques. En épidémiologie, les réseaux ont été utilisés pour modéliser des situations épidémiques dans des populations hétérogènes.[7, 41, 48, 83, 91] Les relations entre les différentes entités du réseau représentent des interactions qui présentent un intérêt au regard de l'épidémie étudiée : un réseau permet la représentation microscopique des liens pouvant conduire à la dissémination de l'agent infectieux entre les individus. Historiquement, Burnet propose, dès 1940, l'utilisation de graphes pour étudier la dissémination d'un pathogène théorique dans une population hétérogène.[18] La difficulté de définir un « contact », et donc un graphe, fait que ce n'est qu'au milieu des années 1980 que ce concept est réellement repris, motivé par l'étude des réseaux sexuels et leur rôle dans la propagation du VIH,[90] récemment découvert à l'époque. Depuis, les réseaux occupent une place croissante en épidémiologie, car ils permettent de représenter de façon fidèle l'hétérogénéité individuelle des « contacts ». Danon et al. proposent une revue des méthodes d'utilisation des réseaux depuis leur (ré)introduction en épidémiologie vers 1985.[34]

##### **1.4.1.1 Définition d'un réseau**

###### **1.4.1.1.1 Définition formelle**

On appelle  $G(V_n, E)$  le réseau (ou graphe) composé d'un ensemble de  $n$  nœuds ( $V_n$ ) et d'un ensemble d'arêtes  $E$ . Les arêtes peuvent être dirigées ou non : dans le second cas, toute arête est considérée comme symétrique. Elles peuvent également être pondérées :

dans le cas d'un réseau de contacts, par les durées de contacts entre individus par exemple.

#### 1.4.1.1.2 Représentation

Un réseau  $G(V_n, E)$  peut être représenté par une matrice d'adjacence ou de liste d'arêtes. On appelle matrice d'adjacence  $A$  relative au graphe  $G(V_n, E)$  la matrice de dimension  $n * n$  où :

$$a_{ij} = \begin{cases} w_{ij} & \text{si une arête relie les noeuds } i \text{ et } j \\ 0 & \text{sinon} \end{cases}$$

Dans un réseau non-pondéré et symétrique,  $A$  est une matrice symétrique où la diagonale est nulle (un nœud n'établit pas de lien avec lui-même) et où une arête prend la valeur 1 lorsqu'elle existe.

#### 1.4.1.2 Attributs

En plus de la donnée relationnelle permettant de construire la topologie du réseau, chaque nœud peut se voir attribuer des caractéristiques le décrivant, par exemple des données sociodémographiques (âge, sexe, localisation) ou épidémiologiques (statut infectieux). Ce sont des attributs nodaux.

#### 1.4.1.3 Mesures sur des réseaux

La description d'un réseau utilise des mesures héritées de la théorie des graphes. Ces mesures décrivent la topologie du réseau, au niveau d'un nœud ou dans son ensemble. En voici une liste non-exhaustive.

##### 1.4.1.3.1 Densité

La densité d'un graphe est le nombre d'arêtes divisé par le nombre total d'interactions possibles (soit  $\frac{n*(n-1)}{2}$  dans un réseau non-dirigé, ou le double dans un réseau dirigé).

##### 1.4.1.3.2 Degré

Le degré d'un nœud correspond à la taille de son voisinage. Dans un graphe non-dirigé, le degré d'un nœud est le nombre de nœuds auquel il est lié par des arêtes. Dans un graphe dirigé, le degré entrant d'un nœud  $i$  est le nombre de nœuds  $j$  ( $\neq i$ ) ayant une arête dirigée vers  $i$  ; le degré sortant est le nombre de nœuds  $j'$  ( $\neq i$ ) vers lesquels  $i$  a une arête sortante. La distribution du degré, souvent notée  $P(k)$  définit la probabilité pour un nœud au hasard d'avoir  $k$  voisins sur le graphe. Cette quantité a été identifiée comme

à l'origine d'une forte disparité dans le potentiel de dissémination épidémiologique d'un individu.[83, 137]

#### 1.4.1.3.3 Distance minimale

La distance entre deux nœuds correspond à la somme des poids des arêtes qui les relient. Quand plusieurs chemins relient deux nœuds, la distance est celle du chemin le plus court. Dans un réseau où toutes les arêtes sont de poids 1, la distance entre deux nœuds  $i$  et  $j$  est le nombre minimum d'intermédiaires qui les relient plus un.

#### 1.4.1.3.4 Diamètre

Le diamètre d'un réseau, noté  $L$ , est la plus grande de ses distances minimales : c'est donc souvent le nombre minimum d'intermédiaires séparant les deux individus les plus éloignés, plus un.

#### 1.4.1.3.5 Centralité

La centralité caractérise la propension d'un nœud à se retrouver lié aux autres nœuds. Plusieurs définitions en existent. Nous retiendrons la centralité d'intermédiaire (« betweenness centrality »). Pour un nœud  $i$ , c'est la proportion des plus courts chemins entre tous les nœuds  $s$  et  $t$  du graphe ( $\sigma_{st}$ ) passant par  $i$  ( $\sigma_{st}(i)$ ) :

$$g(i) = \sum_{s \neq i \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

#### 1.4.1.3.6 Coefficient de clustering

Une mesure naturelle du clustering est:

$$\phi = \frac{3 * \text{Nombre de triangles}}{\text{Nombre de triplets connectés}}$$

Cette donnée capture la propension de deux individus partageant un voisin commun à être eux-mêmes reliés dans le réseau. Dans les réseaux à fort clustering, les individus tendent à se regrouper en communauté de forte densité, ayant peu de connexions avec les autres communautés, et la taille finale de l'épidémie est diminuée,[82] ainsi que le taux de croissance du pathogène (donc, implicitement, le nombre de reproduction).[102]

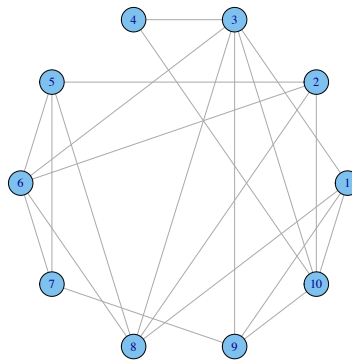
### 1.4.2 Les différents types de réseaux

On distingue dans la littérature trois grandes familles de réseaux.

#### 1.4.2.1 Réseaux aléatoires

Un réseau est dit aléatoire (ou de Bernouilli) lorsqu'il suit le modèle proposé par Erdős et Rényi : chaque arête a une probabilité  $0 \leq p \leq 1$  d'exister et est indépendante des autres arêtes.[47] La distribution du degré est donc binomiale (Figure 8) et l'hétérogénéité de cette distribution est faible. Par construction, ces réseaux ont un coefficient de clustering égal à  $\phi_{random} = p$ . Le diamètre est de l'ordre de  $L_{random} = \frac{\ln(n)}{\ln(np)}$ .

Cette famille de réseaux a longtemps servi de base pour la réflexion sur les systèmes complexes, en faisant l'hypothèse que les réseaux observés étaient aléatoires. Cette théorie s'oppose par nature au concept « d'organisation » : l'intuition veut que les réseaux observés soient structurés et donc ne soient pas l'issue d'une expérience aléatoire.



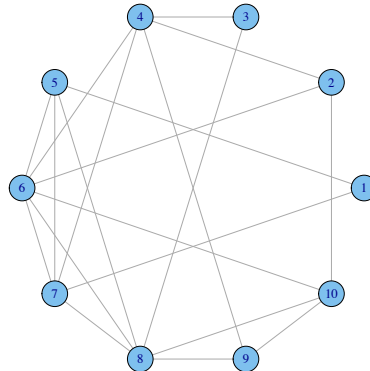
**Figure 8. Réseau aléatoire suivant le modèle d'Erdős-Rényi.** La probabilité d'occurrence d'une arête est  $p = 0.4$ .

#### 1.4.2.2 Réseaux petit-monde (« small-world networks »)

Le phénomène de « petit-monde » permet de relier plusieurs communautés distinctes par des intermédiaires servant de « hub » (Figure 9). Ces réseaux ont été décrits par Watts & Strogatz et se distinguent des réseaux aléatoires par l'introduction du concept d'homophilie : les nœuds établissent des liens de façon préférentielle, et créent donc des communautés à fort clustering, mais peu reliées entre elles. Watts & Strogatz proposent un algorithme pour générer de tels graphes à partir d'une lattice.[152] Les individus reliant ces communautés permettent d'obtenir un diamètre relativement faible, proportionnel à  $\ln(n)$ .

Un réseau est dit « petit-monde » quand :

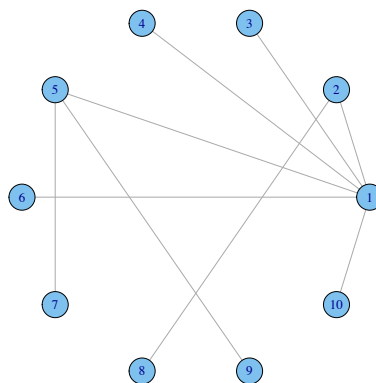
$$\begin{cases} \phi \gg \phi_{random} \\ L \ll L_{random} \end{cases}$$



**Figure 9. Réseau "petit-monde" au sens de Watts & Strogatz.** La probabilité de réarrangement est  $p = 0.4$  pour le voisinage 2-hop sur la lattice de départ.

#### 1.4.2.3 Réseaux invariants d'échelle (« scale-free networks »)

Dans des réseaux invariants d'échelle, la distribution du degré n'est pas aléatoire, mais suit une loi puissance :  $P(k) \propto k^{-\gamma}$ , où  $\gamma$  est un réel positif (Figure 10).[8] Ce modèle permet d'envisager la croissance exponentielle d'un réseau dans le temps, par l'ajout de nœuds, qui seront reliés préférentiellement à d'autres nœuds ayant un degré élevé. Cette probabilité peut être comparée à la mise en place d'arêtes avec des nœuds « populaires », *i.e.* ayant déjà beaucoup de liens. Ils ont donc également des propriétés « petit-monde ». Le réseau des liens sur Internet est par exemple invariant d'échelle[2]. Ces réseaux ont souvent la particularité d'avoir un diamètre très faible, proportionnel à  $\ln(\ln(n))$ . [26]

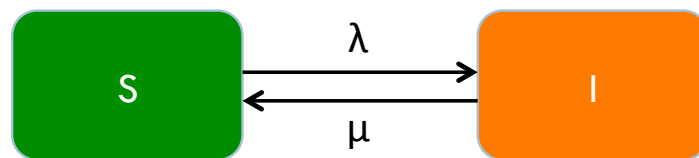


**Figure 10. Réseau invariant d'échelle.** Le paramètre de puissance représente une forme d'homophilie et vaut ici  $\gamma = 2$ .

Nous allons maintenant discuter de l'utilisation des réseaux de contacts dans la modélisation épidémiologique de la propagation d'une bactérie, et des relations entre leur structure et la possibilité pour une épidémie de démarrer.

### 1.5 Paramètres épidémiologiques et structures de contacts

Dans le cadre d'une épidémie émergente, de nombreuses méthodes permettent d'estimer le taux de reproduction de base  $R_0$  (ou une quantité s'en approchant), à partir de l'observation de l'incidence et de paramètres liés à l'histoire naturelle du pathogène (temps de génération, taux de récupération, existence ou non d'une période de latence).[13, 22, 55, 114, 146, 147] Ce taux de reproduction peut aussi être déterminé de façon temps-dépendante, afin de suivre son évolution alors que la structure des contacts change, par exemple par l'implémentation de mesures de contrôle. La détermination de  $R_0$ , ou d'un paramètre s'en approchant, permet de prédire si l'épidémie devrait s'éteindre, ou continuer d'exploser. Nous nous intéressons dans la suite à un modèle de transmission de type S-I-S, qui correspond bien au contexte de colonisation par une bactérie : à la différence du modèle S-I-R, les individus sortant du compartiment infectieux (colonisés) redeviennent immédiatement susceptibles (Figure 11).



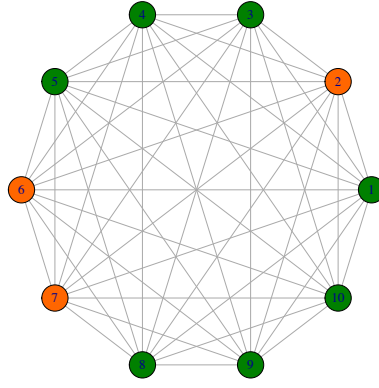
**Figure 11. Modèle Susceptible (S), Infectieux (I).** Les paramètres  $\lambda$  et  $\mu$  représentent les probabilités de transition entre les deux compartiments.

Nous allons présenter la nature du seuil épidémique dans différentes configurations de la structure de contacts. Cette section (et notamment les deux paragraphes qui suivent) s'inspire des ouvrages de Anderson & May,[5] Keeling & Rohani[84] et Barrat, Barthélemy & Vespignani[11].

#### 1.5.1 L'hypothèse simple de transmissibilité densité-dépendante

Dans un modèle S-I-S, l'hypothèse la plus simple est encore celle de mélange homogène complet : tous les individus sont supposés en contact les uns avec les autres Figure 12.





**Figure 12. Représentation d'une population en situation de mélange complet et en présence d'un pathogène.** Les nœuds en vert sont susceptibles, et ceux en orange sont infectieux.

L'équation différentielle qui donne la variation du nombre d'individus infectieux S-I-S est :

$$\frac{dI}{dt} = \lambda SI - \mu I$$

où  $\lambda$  représente le taux de contact efficace et  $\mu$  l'inverse de la période infectieuse. Pour que cette quantité soit non-nulle à la situation initiale, on retrouve le théorème du seuil :

$$\left(\frac{\lambda S_0}{\mu}\right) \geq 1$$

Lorsque l'ensemble de la population est susceptible à l'état initial (*i.e.*  $S_0 \approx n$ ), le seuil épidémique critique devient :

$$\left(\frac{\lambda}{\mu}\right)_c = \frac{1}{n}$$

Cette notion de seuil implique que l'épidémie s'éteindra presque certainement pour les valeurs  $(\lambda, \mu)$  ne satisfaisant pas la condition précédente.

### 1.5.2 Réseaux aléatoires et modèle S-I-S

L'hypothèse la plus minimaliste sur la structure des contacts est celle d'un réseau aléatoire, où la distribution du degré est binomiale (voir paragraphe 1.4.2.1, chaque arête dans un réseau de taille  $n$  existe avec une probabilité  $p$ ) : en moyenne, les individus ont tous le même nombre de voisins. Soit  $K = E(P(k)) = (n - 1) * p$  cette quantité.

L'équation différentielle qui gouverne le modèle est donc :

$$\frac{dI}{dt} = \frac{K\lambda SI}{n} - \mu I$$

De nouveau, pour que l'épidémie démarre dans une population presque entièrement susceptible (*i.e.*  $S_0 \approx n$ ), il faut réaliser la condition :

$$\left(\frac{K\lambda}{\mu}\right) \geq \frac{n}{S_0} \approx 1$$

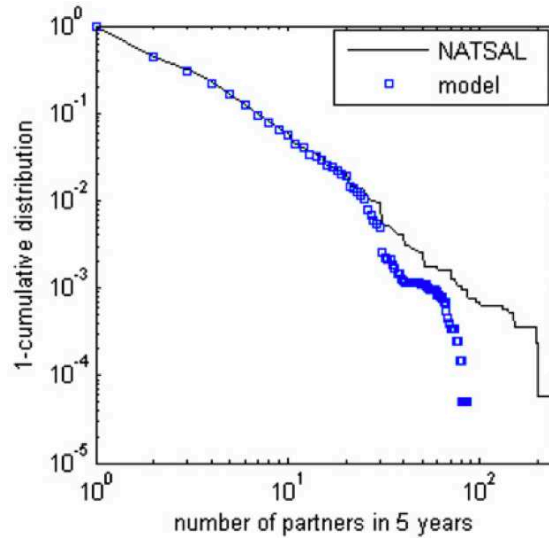
Dans le cas d'une transmission fréquence-dépendante, avec un réseau de contact où chaque nœud est relié en moyenne à  $K$  autres, le seuil épidémique est donc :

$$\left(\frac{\lambda}{\mu}\right)_c = \frac{1}{K}$$

Toutefois, cette hypothèse de réseau de contact est encore peu réaliste au regard des réseaux observés : l'effet « petit-monde » n'est pas pris en compte, ni l'hétérogénéité individuelle.

### 1.5.3 Cas d'une distribution quelconque du degré

La distribution du degré  $P(k)$  peut être obtenue de façon empirique : un scénario imaginable est que le réseau de contact n'est pas observé directement, mais que les individus d'une population rapportent le nombre de personnes avec qui elles ont été en contact. La Figure 13 ci-dessous présente la distribution du degré dans le cas d'un réseau de contacts sexuels[124] : on remarquera que cette distribution suit approximativement une loi puissance, caractérisant un réseau de type « invariant d'échelle ».



**Figure 13. Distribution du nombre de partenaires sexuels (contacts) dans une population observée (trait plein) ou modélisée (points bleus), sur une échelle log-log.** Figure adaptée de Robinson et al. (2012)

Plusieurs études[101, 112, 115] ont montré que la transition entre le régime d'extinction et le démarrage de l'épidémie dépendait de la quantité :

$$\left(\frac{\lambda}{\mu}\right)_c = \frac{\langle k \rangle}{\langle k^2 \rangle}$$

Dans cette équation,  $\langle k \rangle$  et  $\langle k^2 \rangle$  sont les moments d'ordre 1 et 2 du degré (moment d'ordre  $n$  :  $\langle k^n \rangle = \sum_i i^n P(X = i)$ ), c'est à dire le degré moyen et le degré carré moyen. En particulier, dans les réseaux invariants d'échelles où  $\gamma < 3$ , le moment ordinaire d'ordre 2,  $\langle k^2 \rangle$ , tend vers l'infini et il découle que :

$$\left(\frac{\lambda}{\mu}\right)_c \xrightarrow{N \rightarrow +\infty} 0$$

En 2010, une autre étude a exploré davantage le cas de ces réseaux, et étendu le calcul du seuil critique pour d'autres valeurs de  $\gamma$ . [19] Dans ces réseaux, le seuil épidémique est d'autant plus faible que la taille de réseau est grande.

#### 1.5.4 Prise en compte de la structure explicite des contacts

Bien que capturant de plus en plus finement l'hétérogénéité de la structure des contacts, la distribution du degré ne permet pas à elle seule d'évaluer d'un point de vue microscopique les différences de risques entre individus lors du déroulement d'une épidémie.

Soit  $A$  la matrice d'adjacence observée. L'évolution de l'épidémie est un processus Markovien, où à chaque pas de temps, chaque nœud est dans un des deux états  $\{S, I\}$ , soit  $2^V$  états possibles. [60, 150] La probabilité pour un nœud  $i$  d'être infectieux au temps  $t+1$  conditionnellement à son état au temps  $t$  est :

$$P(i, t + 1) = 1 - [1 - (1 - \mu) * P(i, t)] * \prod_{j \neq i} [1 - \lambda A_{ji} P(j, t)]$$

Dans cette équation, le premier terme correspond à la probabilité de rester infectieux si le nœud l'était déjà au temps  $t$ , et le second terme correspond à la probabilité de ne pas être infecté par un seul de ses voisins. Dans ce cas, le seuil épidémique correspond à :

$$\left(\frac{\lambda}{\mu}\right)_c = \frac{1}{\rho(A)}$$

Où  $\rho(A)$  est la plus grande valeur propre de la matrice d'adjacence. [16, 24, 60, 150]

\* \* \*

Nous l'avons vu dans ce chapitre, les modèles épidémiologiques ont évolué depuis leur introduction, afin de prendre en compte des structures de contacts plus réalistes, et donc de représenter plus fidèlement la dissémination d'un pathogène dans une population aux contacts très hétérogènes. Nous n'avons toutefois pas abordé le cas des réseaux dynamiques, qui fera l'objet d'une autre section. Après avoir passé en revue l'état de l'art dans le domaine de l'épidémiologie des réseaux, nous allons maintenant nous intéresser aux données de l'étude I-Bird, où le réseau de contact et le portage d'une bactérie ont été conjointement enregistrés pendant 6 mois. Cette étude s'inscrit dans le cadre de la lutte contre la dissémination des souches nosocomiales antibiorésistantes.

## **2 Chapitre 2 : Épidémiologie à l'ère du numérique**

Dans ce chapitre, nous allons voir que les approches traditionnelles en épidémiologie peuvent bénéficier de la collecte de nouvelles données individuelles, rendues faciles par l'avancée des technologies de l'information et des communications. Nous verrons dans un premier temps les données ayant déjà été collectées dans ce contexte, et nous détaillerons ensuite le projet I-Bird (Individual-based Investigation of Resistance Dissemination), dont les données font l'objet des articles présentés dans cette thèse.

### **2.1 Quel type de données de contacts ?**

Par le passé, les réseaux servant de support à la dissémination de maladies infectieuses étaient beaucoup plus restreints dans le temps comme dans l'espace, les moyens de transport à grande échelle étant moins développés.[77, 113] Historiquement, l'évaluation de ces réseaux a utilisé les cartes de moyens de transport (routes, voies ferrées), et plus récemment des données issues du transport aérien civil.[6] Ces données, liées aux déplacements des individus, permettent d'informer des modèles individus-centrés ou de métapopulations.[64] A l'échelle d'un pays, les variations d'incidence de certaines bactéries nosocomiales ont été expliquées par le réseau généré par les transferts de patients entre hôpitaux au Royaume-Uni.[39] Enfin, les interactions entre individus ont souvent été évaluées par des questionnaires, où les personnes interrogées devaient rapporter les contacts qu'elles avaient eu au cours d'une certaine période, évaluer leurs durées, et éventuellement fournir des informations complémentaires sur le lieu ou le type de contact (physique ou de proximité).[106, 108]

#### **2.1.1 Questionnaires, carnets de bord, investigateurs dédiés**

L'approche la plus intuitive pour reconstruire le réseau de contacts entre des individus consiste à les interroger afin d'obtenir des données égocentrées.[106] Dans ce cas, les informations collectées sont non-symétriques : chaque individu rapporte les contacts dont il se souvient. Une alternative possible est de fournir à chaque participant un carnet de bord, dans lequel il doit consigner les contacts qu'il a, ainsi que des données démographiques à propos des personnes rencontrées.[45] Ce mode de collecte est celui utilisé par l'étude POLYMOD.[108] Il présente l'avantage de permettre de retranscrire les contacts en temps (presque) réel, et peut-être de limiter l'omission de certains

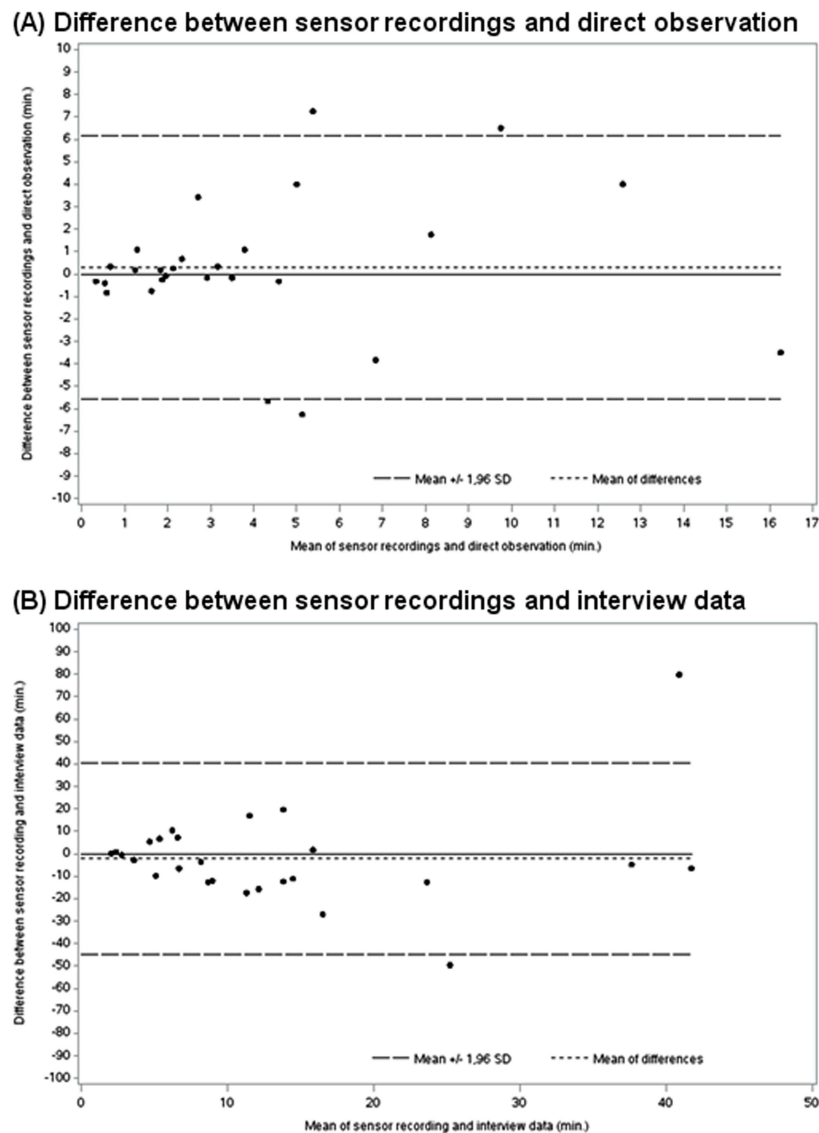
contacts. Dans les deux cas, l'absence de symétrie permet d'imputer une partie des données manquantes selon des règles simples : si un individu rapporte une interaction avec un autre, alors la même interaction devrait être rapportée par son contact. Smieszek et al.[132] ont montré que ces données étaient affectées par plusieurs biais : les femmes tendent à rapporter des données plus exhaustives que les hommes, et les participants sous-évaluent les contacts de faible durée. A l'inverse, les contacts longs sont globalement bien rapportés.

Enfin, même si cette technique requiert des moyens humains plus importants, certaines études font appel à des investigateurs dédiés, chargés de suivre les participants et de noter tous les contacts qu'ils ont.[74, 99] Cette méthode, coûteuse en temps et en ressources, est surtout utilisée pour évaluer l'efficacité d'autres moyens de rapport des contacts.

### **2.1.2 Collecte massive et automatisée par des capteurs autonomes**

La révolution de la technologie de l'information permet le recueil des contacts entre individus avec une précision sans précédent. La localisation des individus peut se faire par leur téléphone portable,[61] des métadonnées informatiques,[129] ou encore des capteurs électroniques[10, 20, 79, 130, 144] offrant une méthode moins intrusive et plus facile à maîtriser. Dans ces études, chaque participant se voit attribuer un capteur électronique de petite taille, capable d'identifier en temps réel les autres capteurs se trouvant à proximité immédiate. Le seuil de détection peut être calibré selon les besoins de l'étude.

En 2012, Lucet et al. montrent que ce type de capteurs permet d'enregistrer avec une grande fidélité les contacts entre patients et personnels soignants dans un service de maladies infectieuses (Figure 14).[99]



**Figure 14. Graphiques de Bland-Altman des contacts enregistrés.** Différence entre les enregistrements des capteurs et les observations directes (A) ou les sondages *a posteriori* (B). Figure reproduite de Lucet et al. (2012)

Une autre étude rapporte peu de congruence pour le croisement de données issues de capteurs similaires avec les contacts rapportés par des questionnaires conduits le lendemain des contacts.[132] L'hypothèse la plus plausible est que les contacts rapportés de façon rétroactive, suite à un questionnaire, omettent les interactions de courte durée ou avec des interlocuteurs inhabituels, qui n'ont pas « marqué » les personnes.

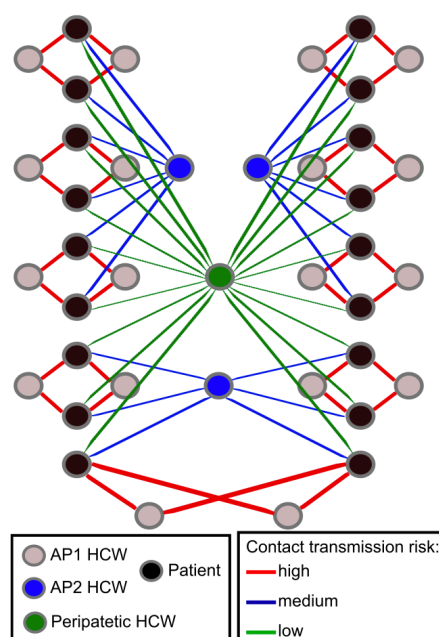
## 2.2 Réseaux de contacts à haute résolution dans la littérature

### 2.2.1 Approche théorique pour caractériser des « super-transmetteurs »

En 2009, Temime et al. ont modélisé la structure des contacts dans un service de réanimation.[137] Par réalisme, les interactions entre patients n'étaient pas envisagées, et trois types de personnels de santé étaient considérés :

- Personnel de santé ayant beaucoup de contacts avec peu de patients (ex : infirmier/ère),
- Personnel de santé ayant peu de contacts avec plus de patients (ex : médecin),
- Personnel de santé ayant un contact quotidien avec tous les patients (ex : radiologue, kiné/ergothérapeute).

Le réseau des interactions ainsi modélisées est rapporté dans la Figure 15. Les interactions entre personnel de santé sont modélisées à partir de données publiées par d'autres équipes, recueillies dans des services d'urgences et sont supposées surévaluer les contacts entre personnel de santé dans le service modélisé.[73, 74]



**Figure 15. Réseau des contacts modélisés dans un service de réanimation.** Les 18 patients (en noir) sont en contacts avec 3 catégories de personnels de santé. Le personnel de santé de type 1 ou 2 (respectivement gris et bleu) est assigné à des sous-groupes de patients, tandis que l'unique personnel de santé de type 3 a des contacts avec tous les patients (vert). La couleur des liens représente le risque de transmission de SARM à chaque contact (low : 3-10% ; medium : 5.5-17% ; high : 8.5-25%). Figure reproduite de Temime et al. (2009)



Ces profils ont été croisés à différents niveaux d'adéquation aux mesures d'hygiène des mains dans le service, afin de simuler la dissémination d'une bactérie nosocomiale (*S. aureus* résistant à la méticilline et entérocoque résistant à la vancomycine) sur ce réseau. Les auteurs identifient les personnels de santé ayant un seul contact avec tous les patients (en vert sur la figure) comme potentiels « super-transmetteurs », causant l'augmentation la plus importante d'incidence lorsque l'adéquation aux mesures d'hygiène diminue. Ces résultats sont en accord avec une autre étude théorique où les médecins qui entraient en contact avec des patients de différents services causaient la plus forte augmentation d'incidence.[140]

Hornbeck et. al identifient, en 2012, des personnels de santé ayant le profil à risque théorisé par Temime et. al en 2009,[137] en observant deux journées de données de contacts dans un service de réanimation.[76]

## **2.2.2 Application pratique**

### **2.2.2.1 Approche observationnelle**

Plusieurs exemples de réseaux à haute résolution temporelle existent dans la littérature. En 2006, le projet OpenBeacon (<http://www.openbeacon.org/>) se lance dans la mise au point de capteurs de proximité autonomes, capables de détecter les autres capteurs situés à moins d'une certaine distance. Dès 2008, des capteurs issus de ce projet sont utilisés pour enregistrer le réseau d'interactions de faible proximité entre les participants de plusieurs conférences.[20] Ces capteurs utilisent une infrastructure fixe pour le stockage des données : les interactions détectées sont envoyées à un récepteur fixe qui se charge de centraliser l'horodatage et le stockage des informations. Dans cette étude, les auteurs identifient un lien supra-linéaire entre le nombre de contacts d'un individu et leur durée : les participants qui ont le plus de contacts ont également les contacts les plus longs. Ils émettent l'hypothèse que ces personnes façonnent les phénomènes de diffusion utilisant le réseau comme substrat, en accord avec de précédentes études théoriques.[5, 103, 115]

### **2.2.2.2 Evaluation rétroactive de mesures de contrôles informées par le réseau des contacts**

En 2009, de nouvelles études enregistrent des réseaux de contacts à haute résolution dans un service de pédiatrie[79] et dans un lycée[130]. Salathé et al., en plus de décrire

un réseau de contact de type « petit-monde » relativement homogène (distribution du degré et des durées de contacts), modélisent la dissémination d'une épidémie de grippe selon un modèle SEIR et testent des stratégies de contrôle naïves ou utilisant les données du réseau.. Les paramètres de leur modèle de transmission sont repris d'autres études décrivant l'histoire naturelle du virus de la grippe.[52, 97] Les approches de vaccination ciblant les nœuds fortement connectés dans le réseau permettent de réduire significativement l'incidence de grippe par rapport à une vaccination aléatoire.

\* \* \*

L'épidémiologie à l'ère du « big data » permet d'utiliser des données très précises pour informer les modèles de transmission sur les interactions entre individus. La limitation de la dissémination de souches multirésistantes en milieu hospitalier est un des grands défis actuels en santé publique. Le projet européen MOSAR (Mastering hOSpital Antimicrobial Resistance, <http://www.mosar-sic.org/>) a proposé divers axes d'études sur la dissémination de l'antibiorésistance. En particulier, son volet I-Bird (Individual-based Investigation of Resistance Dissemination ; MOSAR WP7 et WP8) cherche à comprendre les liens entre structures de contacts et dissémination de certaines bactéries nosocomiales.

### **2.3 Le projet I-Bird**

Le projet I-Bird se concentre sur la dissémination de *S. aureus*, responsable d'une grand part des infections nosocomiales.

#### **2.3.1 Le poids des infections nosocomiales à *S. aureus***

Les premières souches de *S. aureus* résistantes à la méticilline (SARM, par opposition aux souches sensibles, dites SASM) ont été observées dans les années,[9, 80] quelques années seulement après l'introduction de cet antibiotique. Les mécanismes de résistance sont connus, et reposent principalement sur l'altération de protéines de surfaces.[71, 141] Le portage de *S. aureus* est fréquent (surtout en milieu hospitalier), souvent multiple, et affecte plusieurs parties du corps (peau, nasopharynx, blessures, emplacement de matériel médical invasif...).[149, 157] Les SARM sont source d'une

morbidité et mortalité accrue par rapport aux SASM,[30, 153] bien que la virulence des SASM et SARM ne semble pas différer significativement.[32, 88, 104] L'augmentation des infections à SARM est le reflet de l'impact croissant des procédures médicales sur des patients présentant de plus en plus de comorbidités : un effet de bord de l'avancée dans le domaine de la médecine.[17] Parmi les infections pouvant découler d'une colonisation par *S. aureus*, nous noterons les infections cutanées, respiratoires ou encore septicémies[93] : aux Etats-Unis, l'incidence annuelle des infections à *S. aureus* a été estimée à 28.4 cas pour 100 000 habitants, soit un taux équivalent aux infections par pneumocoques et plus élevé que pour les streptocoques.[94] En Europe, *S. aureus* est également parmi les microorganismes les plus souvent à l'origine d'infections nosocomiales.[119]

La rapidité d'apparition de souches résistantes en fait un défi majeur de santé publique : la réduction de l'incidence des infections passe par un meilleur contrôle de leur transmission. *S. aureus* est principalement disséminé par l'intermédiaire d'individus colonisés de façon transitoire (notamment le personnel de santé), à l'interface entre un réservoir (patient colonisé) et un autre individu susceptible d'être colonisé.

### **2.3.2 Stratégies de contrôle de la transmission de bactéries nosocomiales**

Le contrôle de la transmission fait intervenir deux stratégies : l'adéquation aux mesures d'hygiène permettant d'éradiquer le pathogène lorsqu'un portage transitoire est possible, et l'organisation des contacts afin de minimiser les risques d'acquisition et donc de dissémination.

#### **2.3.2.1 Contrôle direct par application des mesures d'hygiène**

La stratégie la plus évidente pour limiter la dissémination est le respect des normes d'hygiène lors des procédures médicales. La revue de Pittet et al. décrit les processus menant à la dissémination de germes nosocomiaux lors d'une mauvaise adéquation à ces mesures.[117] Améliorer le respect des procédures d'hygiène fait l'objet de recommandations de l'ensemble de la communauté scientifique.[155]

#### **2.3.2.2 Contrôle indirect par l'étude des interactions interindividuelles**

Nous avons vu dans le précédent chapitre que les réseaux de contacts pouvaient servir à informer des modèles épidémiologiques. Les interactions entre patients et personnels

de santé dans un hôpital sont donc susceptibles d'influencer la dissémination des bactéries nosocomiales.

### **2.3.3 Collecte des données I-Bird et exploitations possibles**

I-Bird est une étude prospective longitudinale réalisée dans l'hôpital maritime de Berck-sur-Mer, entre mai et novembre 2009, afin d'évaluer l'influence du réseau de contact sur le portage individuel de *S. aureus*. Dans le cadre de cette étude, les patients et le personnel de santé ont accepté de porter un capteur de proximité pendant toute la durée passée à l'hôpital, ainsi que de se soumettre à un écouvillonnage nasal hebdomadaire afin de détecter le portage de *S. aureus* (taux de participation : 90.1%, soit 592 individus). D'autres prélèvements étaient effectués de façon non systématique (écouvillonnage rectal chez les patients, prélèvements sur les escarres et sur le matériel invasif). Nous nous limiterons dans cette thèse à l'exploitation des données allant du 30 juin au 25 octobre 2009, les deux premiers mois d'enregistrement ayant servi de phase pilote.

#### **2.3.3.1 Réseau de contact à haute résolution**

Chaque participant devait porter un capteur à tout moment lorsqu'il était dans l'hôpital. Ce capteur est capable de détecter, toutes les 30 secondes, les autres capteurs se trouvant à proximité (en ligne de vue à moins de 1.5 m). Chaque interaction est enregistrée dans la mémoire du capteur et horodatée. L'enregistrement a été interrompu 4 fois au cours de l'étude, 2 jours à chaque fois, afin de remplacer l'intégralité des batteries des capteurs. L'ensemble des données collectées fournit un réseau de contacts où les nœuds sont les individus (patients et personnels de santé) et où les arêtes représentent une interaction sociale d'au moins 30 secondes entre deux nœuds.

#### **2.3.3.2 Suivi longitudinal du portage de *S. aureus***

Les participants se soumettaient chaque semaine à un écouvillonnage nasal : tous les prélèvements étaient mis en culture puis incubés 48 h afin de détecter la présence de *S. aureus*. Lorsque des souches étaient présentes, elles étaient systématiquement génotypées, afin de déterminer le *spa* type de la souche (identifiant permettant de classer les souches selon les répétitions de certaines séquences du gène codant pour la protéine A).[57] Chaque souche était également mise en culture afin de déterminer son

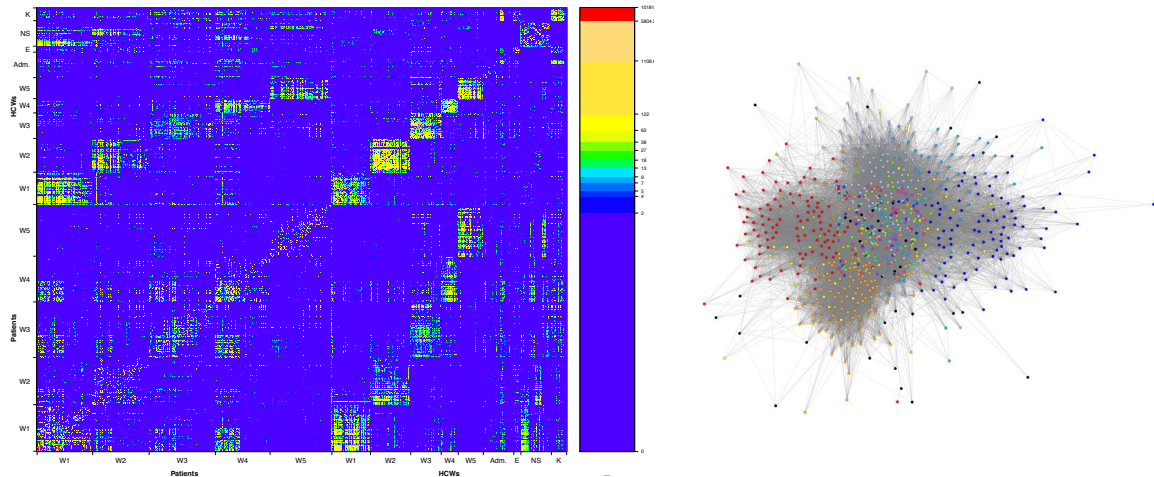
profil de résistance à 20 antibiotiques (dont la méticilline). L'ensemble de la procédure est décrit plus en détail dans le S1 Text de l'article présenté au paragraphe 3.3.2.

### 2.3.4 Graphes de contacts « petit-monde » au sein de l'hôpital

L'ensemble des interactions recueillies permet de construire un réseau hautement dynamique, constitué de 592 nœuds et 2 671 832 arêtes.

#### 2.3.4.1 Réseau complet agrégé sur l'ensemble de l'étude

Le réseau composé de l'ensemble de ces interactions est dense (27% des interactions possibles) et fait apparaître 5 communautés correspondant aux 5 services de l'hôpital (algorithme de Louvain[15]). Ces communautés sont très denses, et reliées aux autres communautés par peu d'individus. Nous observons une assortativité par service, et même par catégorie professionnelle au sein de l'hôpital (Figure 16).

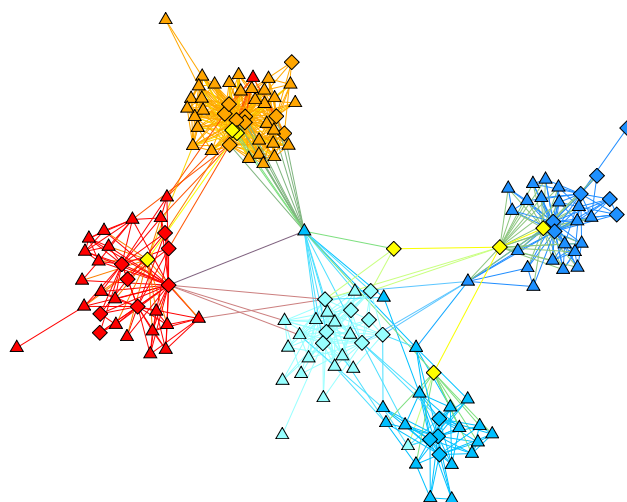


**Figure 16. Les contacts dans l'étude I-Bird. (Gauche)** Matrice des contacts par service et par catégories de personnels de santé. Les contacts se produisant avec une fréquence faible sont en bleu, ceux se produisant assez souvent en vert, et très souvent en rouge. **(Droite)** Réseau agrégé des contacts de l'étude I-Bird, contenant tous les liens indépendamment de leur date d'occurrence. Les couleurs correspondent aux bâtiments/services ou à la catégorie de personnel de santé (orange/rouge : W1 et W2 ; 3 nuances de bleu : W3 à W5 ; jaune : personnel de nuit ; vert : ergothérapeute ; rose : kinésithérapeute ; noir : administratif/autre). L'algorithme de placement des nœuds n'utilise pas d'information sur le service.

#### 2.3.4.2 Propriétés des réseaux discrétisés par jour

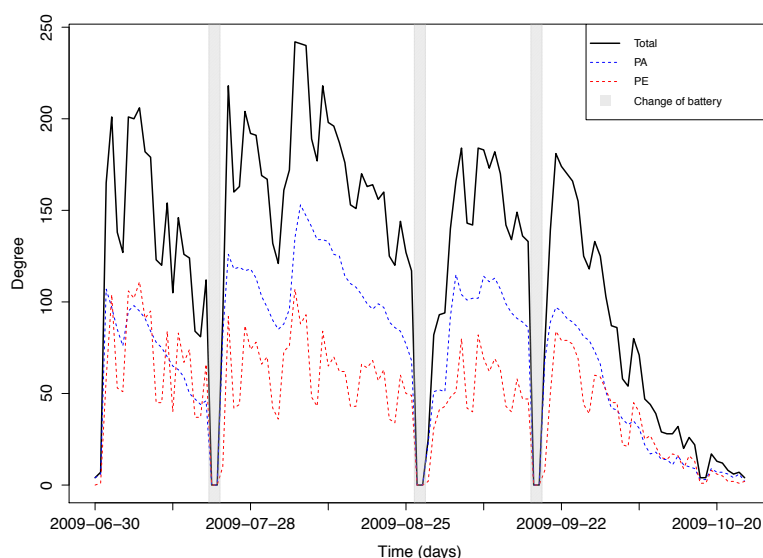
Dans la suite, nous considérerons la collection de réseaux obtenus par agrégation quotidienne (de minuit à minuit) des arêtes du réseau. Ainsi, si deux individus  $i$  et  $j$  ont eu  $N_{ij}$  contacts de durée  $D_{ij} = \{d_{ij}^{(1)}, d_{ij}^{(2)}, \dots, d_{ij}^{(N)}\}$  un jour donné, ces contacts seront agrégés en une seule arête de durée  $\Delta_{ij} = \sum_{k=1}^N d_{ij}^{(k)}$ . Soit  $G(V(t), E(t), t)$  la collection de l'ensemble des réseaux de contacts quotidiens ainsi définis. Ces réseaux présentent des

propriétés similaires à celles du réseau complet, notamment avec l'apparition encore plus nette de 5 communautés (Figure 17).



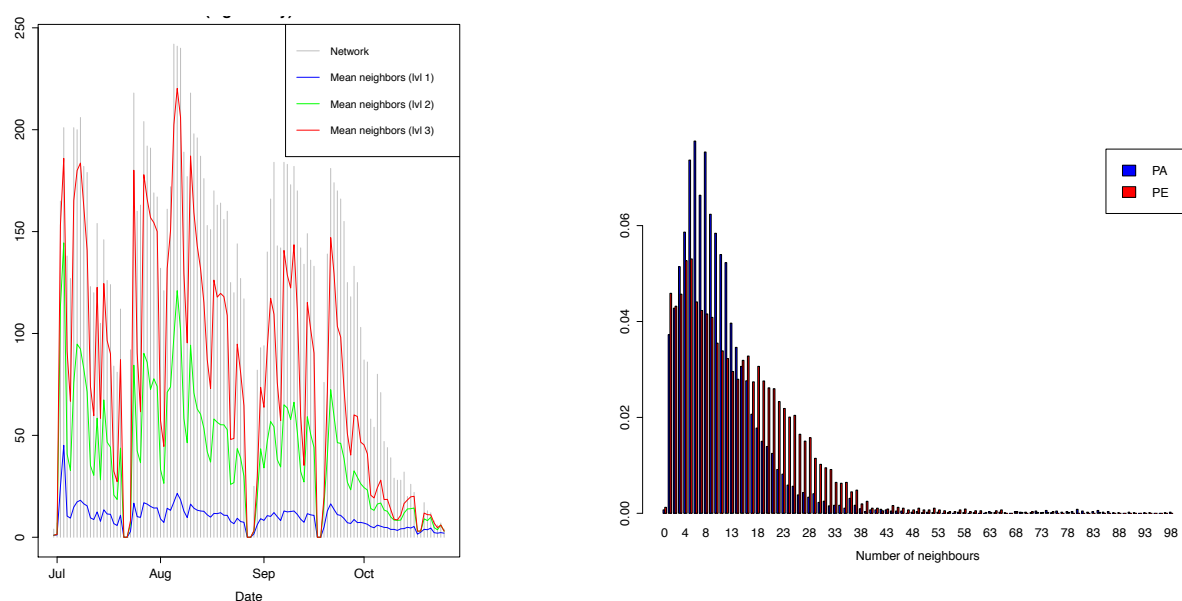
**Figure 17. Contacts de l'étude I-Bird au cours d'une journée typique.** Les patients sont représentés par des triangles et le personnel de santé par des losanges. Les couleurs représentent le service de rattachement ou la catégorie de personnel de santé (orange/rouge : W1 et W2 ; 3 nuances de bleu : W3 à W5 ; jaune : personnel de nuit). La couleur d'un lien est le mélange des couleurs des deux nœuds impliqués. L'algorithme de placement des nœuds n'utilise pas d'information sur le service.

L'analyse de  $G(V(t), E(t), t)$  fait ressortir des propriétés « petit-monde » qui génèrent des graphes souvent connexes. La médiane du nombre d'individus actifs chaque jour est de 139 (2 au minimum, 243 au maximum, IQR : [82-173]). La Figure 18 présente la taille du réseau chaque jour ainsi que la répartition entre patients et personnels de santé.



**Figure 18. Composition du réseau des contacts de l'étude I-Bird au cours du temps.** Le nombre total de nœuds actifs dans le réseau est représenté en trait plein. Les pointillés bleus et rouges correspondent respectivement au nombre de patients et de personnels de santé. Les dates de changement de batteries sont indiquées par des zones grisées.

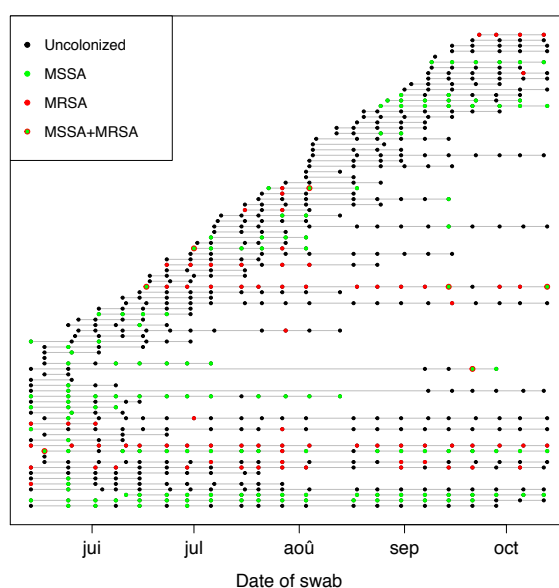
Au cours des 111 jours d'enregistrements disponibles, la percolation est observée 80 fois, et les journées où le graphe n'est pas connexe correspondent aux périodes proches des changements de batteries et aux derniers jours d'enregistrement, où les individus quittant l'étude n'étaient pas remplacés par de nouveaux participants. Le coefficient de clustering est relativement élevé, avec une valeur médiane de 0.39 (IQR : [0.37-0.43]). Dans le cas d'un réseau aléatoire, le coefficient de clustering est proche de la densité du réseau, puisque chaque lien se produit indépendamment des autres liens. Le diamètre médian du réseau est 6 (IQR : [5-7]), beaucoup plus petit que le nombre d'individus actifs. En définissant le « voisinage  $n$ -hop » d'un individu comme l'ensemble des autres individus atteignables sur le réseau par au plus  $n-1$  intermédiaires, nous pouvons illustrer les propriétés « petit-monde » de ces réseaux : la Figure 19 présente, chaque jour, la taille moyenne du voisinage 1-hop (bleu), 2-hop (vert) et 3-hop (rouge) d'un individu dans le réseau. Nous remarquons que le voisinage 3-hop contient presque l'intégralité des nœuds du réseau (ce qui correspond à un diamètre approximativement égal à 6). Enfin, la distribution individuelle du degré permet d'observer que peu d'individus ont beaucoup de contacts, aussi bien chez les patients que chez les personnels de santé.



**Figure 19. Nombre des voisins des participants de l'étude I-Bird. (Gauche)** Taille moyenne du voisinage d'un participant en fonction de la date. Le voisinage direct est représenté en bleu, le voisinage 2-hop en vert et 3-hop en rouge. Les barres grises représentent le nombre total de capteurs actifs chaque jour. **(Droite)** Distribution normalisée du nombre de contacts directe d'un patient (bleu) et d'un personnel de santé (rouge) au cours de l'étude.

### 2.3.5 Suivi des colonisations par *S. aureus*

Parallèlement à l'enregistrement du réseau de contacts, les participants subissaient chaque semaine un écouvillonnage nasal afin de détecter les épisodes de colonisation par *S. aureus*. L'utilisation du *spa* type et du profil de résistance antibiotique de chaque souche fournit une identité fine pour chaque isolat. Un extrait de ces données est présenté dans la Figure 20 : chaque ligne représente un individu, et chaque point un prélèvement. Nous sommes donc en mesure de déterminer des trajectoires longitudinales de portage.



**Figure 20. Colonisation par *S. aureus* dans l'étude I-Bird.** Chaque ligne représente la trajectoire de colonisation d'un participant, et chaque point un prélèvement. La couleur du point indique le statut de colonisation (noir : non colonisé ; vert : colonisation par SASM ; rouge : colonisation par SARM ; vert entouré rouge : colonisation SASM et SARM). Les données présentées ici sont celles des patients et personnel de santé d'un seul service.

\* \* \*

Ce double recueil de données permet non seulement d'étudier les interactions interindividuelles dans un hôpital de soins de suite, mais également d'identifier des événements de transmission de *S. aureus*, lors de la survenue d'épisode d'incidence chez un individu : les modalités de transmission (principalement par contact) font que le réseau des contacts mesurés doit pouvoir expliquer l'incidence observée. Nous nous intéresserons dans le chapitre suivant à tester cette hypothèse.



### **3 Chapitre 3 : Utilisation du réseau des contacts observés pour l'identification de chaînes de transmission de *S. aureus* en milieu hospitalier**

Nous avons vu dans les précédents chapitres que les réseaux de contacts peuvent être utilisés pour réaliser des simulations individu-centrées, où la transmission d'un pathogène modélisé se fait le long des arêtes du réseau. Si le réseau observé est bien le substrat sur lequel se déroule la transmission, il doit capturer le « chemin » suivi par le pathogène, depuis un individu porteur jusqu'à un autre individu nouvellement infecté. Traditionnellement, l'identification d'événements de transmission est réalisée rétrospectivement, par l'identification des personnes avec qui un cas incident a été en contact. Dans ce chapitre, nous utilisons un réseau de contact pour identifier des chemins possibles de transmission. En particulier, nous nous intéresserons aux conséquences que peuvent avoir des résolutions temporelles différentes pour les données de portage et de contacts.

#### **3.1 Stratégies d'identification de chaînes de transmission**

##### **3.1.1 Traçage d'infections (« infection tracing »)**

Le traçage des infections établit *a posteriori* un arbre de transmission cohérent avec les différents épisodes d'incidence.[72, 83, 121] Chaque individu est relié à son infecteur et aux personnes qu'elle a ensuite infectées sur un graphe : chaque arête représente donc des événements de transmission. Ce type de graphe ne souffre pas de la définition (délicate) d'un « contact ». Pour établir un tel graphe, il est nécessaire de documenter l'évolution naturelle du pathogène (temps de génération, période infectieuse). Les voies de transmission identifiées sont le plus souvent basées sur des suppositions spatiales et temporelles (présence d'un individu source et d'un individu cible au même endroit, au même moment).

##### **3.1.2 Traçage des contacts (« contact-tracing »)**

Le traçage des contacts vise à identifier toutes les chaînes de transmission possibles partant d'un individu source. La même recherche peut ensuite être étendue aux contacts à risque, et à leurs propres contacts, etc.[59, 90, 91, 121]. La définition d'un « contact »

est liée au pathogène : il s'agit d'une interaction entre deux personnes qui peut amener à un événement de transmission. D'abord limitée à l'étude des réseaux de contacts sexuels, plus faciles à définir,[38, 90, 105, 154] cette approche a été généralisée aux maladies transmissibles par microparticules aéroportées et fait appel aux contacts sociaux,[21, 67, 101] recueillis de différentes façons (voir section 2.1). Gardy et al. ont même investigué un réseau de contacts (obtenus par questionnaire rétrospectif des patients) croisé à des données génotypiques lors d'une épidémie de tuberculose afin de reconstruire un arbre épidémique plausible.[58]

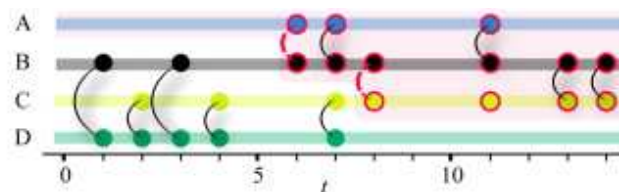
### 3.2 Identification de chaînes de transmission dans un réseau dynamique

L'utilisation d'un réseau statique permet d'identifier rapidement un chemin allant d'un infecteur à un infecté. Idéalement, les infecteurs sont en contact direct avec les personnes qu'ils ont infectées, et il existe une arête reliant ces deux personnes dans le réseau. Toutefois, cela suppose que la maladie se propage rapidement par rapport à la dynamique de formation/dissolution des liens.[43] Nous discutons dans la suite des conséquences de l'agrégation dans un réseau dynamique.

Cette section a pour but d'introduire les outils de calculs utilisés dans l'article présenté au paragraphe 3.3.2.

#### 3.2.1 Agrégation

L'agrégation d'un réseau dynamique en un seul réseau statique soulève le problème de la cohérence temporelle, en faisant apparaître des chemins de propagation qui auraient été impossibles étant donné l'ordre d'apparition et de disparition des liens.[75] Cette problématique est illustrée dans la Figure 21.



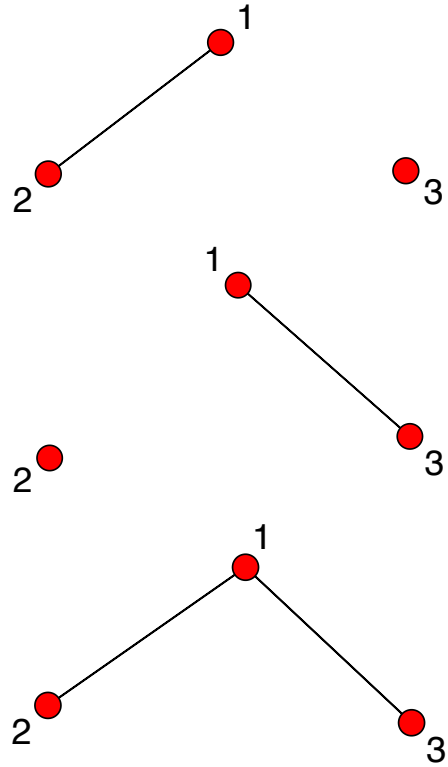
**Figure 21. Représentation temps-dépendante des liens entre 4 nœuds dans un réseau.** A chaque unité de temps, les liens peuvent perdurer, se réarranger ou se défaire. Figure adaptée de Holme et al. (2012)

Considérons un réseau dynamique sur 2 pas de temps et à 3 nœuds, défini par une suite de 2 matrices d'adjacence  $A_{ex}^{(t)}$ ,  $t \in \{1,2\}$ , tel que présenté dans la Figure 22.

$$A_{ex}^{(1)} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$A_{ex}^{(2)} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

$$A_{ex}^* = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$



**Figure 22. Matrices d'adjacence, graphes associés et temporalité.** Les deux premiers graphes représentent une succession de liens (d'abord 2-1, ensuite 1-3). Le troisième graphe représente l'agrégation naïve des deux précédents.

En tenant compte de l'ordre d'occurrence des arêtes, dans ce réseau, un chemin existe entre les nœuds 2 et 3, par l'intermédiaire du nœud 1 : le nœud 2 pourrait infecter le nœud 1, qui à son tour infecterait le nœud 3. Aucun chemin respectant cette temporalité n'existe entre le nœud 3 et le nœud 2 : l'agrégation de plusieurs couches (représentant la dimension du temps) fait apparaître des chemins n'ayant pas lieu d'être, comme dans le réseau  $A_{ex}^*$  où 2 et 3 sont liés *via* 1, indépendamment de l'ordre de survenue des liens 1-2 et 1-3. Plusieurs algorithmes existent pour extraire les plus courts chemins dans des graphes statiques, notamment ceux de Dijkstra[37] ou de Bellman-Ford[12, 54].

### 3.2.2 Chemins temporellement cohérents

#### 3.2.2.1 Existence de tels chemins

Dans l'étude I-Bird, nous disposons de données de portage hebdomadaire de *S. aureus*. Nous étudions la collection de réseaux agrégés par jour et paires, comme précédemment évoqué, avec  $A^{(i)}$  la matrice d'adjacence associée au réseau des contacts du jour  $i$ . La détection d'une nouvelle souche chez un individu laisse à penser qu'il doit exister un chemin le reliant à un individu ayant porté la même souche la semaine précédente. Lorsqu'un tel individu peut être identifié, nous cherchons donc à extraire tous les

chemins partant de cet individu, afin de voir s'il peut être relié de façon temporellement cohérente au cas incident. Soit  $d$  la date de détection de l'incidence du cas index et  $d+\tau$  la date de détection du nouveau cas incident.

#### 3.2.2.1.1 Cas particulier : l'information se propage d'une arête à chaque pas de temps

Notons que dans le cas plus simple où l'information qui transite *via* les arêtes du graphe (*i.e.* ici, le pathogène) se propage exactement à la même vitesse que la formation des arêtes, la matrice :

$$P = \sum_{k=d}^{d+\tau} \left( \sum_{j=k}^{d+\tau} \prod_{i=d}^j A^{(i)} \right)$$

a pour valeurs le nombre total de chemins possibles ralliant deux individus : l'élément  $p_{ij}$  correspond au nombre de chemins partant de  $i$  et atteignant  $j$  pour tous les couples de pas de temps  $d_1$  et  $d_2$  dans l'intervalle  $[d ; d+\tau]$  (avec  $d_1 < d_2$ ). Plus précisément, chaque matrice de la somme principale (indice  $k$ ) contient le nombre de chemins, commençant au moins au temps  $k$  et finissant au temps  $d+\tau$ , qui relient les individus les uns aux autres. Par exemple, pour une série à trois pas de temps :

$$p_{ij} = \left( a_{ij}^{(1)} + a_{ij}^{(1)} a_{ij}^{(2)} + a_{ij}^{(1)} a_{ij}^{(2)} a_{ij}^{(3)} \right) + \left( a_{ij}^{(2)} a_{ij}^{(3)} + a_{ij}^{(3)} \right) + \left( a_{ij}^{(3)} \right)$$

Toutefois, cette approche n'est pas directement applicable aux données de I-Bird, car elle suppose qu'à chaque pas de temps, le pathogène ne peut effectuer qu'un « saut » depuis sa source sur le réseau, ce qui n'est pas réaliste au regard de l'agrégation par jour déjà menée.

#### 3.2.2.1.2 Cas général : l'information se propage sur un nombre variable d'arêtes à chaque pas de temps

Dans le cas plus général, nous recherchons l'existence d'un chemin reliant deux nœuds, en ayant la possibilité de réaliser plusieurs déplacements le long des arrêts à chaque pas de temps. Ce concept fait intervenir les matrices d'accessibilité.[4, 96, 111] La matrice :

$$\mathcal{A}^{(i)} = \sum_{j=1}^N (A^{(i)})^j$$

contient tous les chemins de longueur 1 à  $N$  (le nombre total de nœuds) réalisables sur le réseau dont la matrice d'adjacence est  $A^{(i)}$ .

Lorsque le réseau est dynamique, la matrice d'accessibilité entre deux dates  $t = d$  et  $t = d + \tau$  est :

$$\mathcal{A} = \sum_{i=d}^{d+\tau} \mathcal{A}^{(i)} = \sum_{i=d}^{d+\tau} \sum_{j=1}^N (A^{(i)})^j$$

En remplaçant les matrices d'adjacence par des matrices d'accessibilité dans l'expression de  $P$ , nous obtenons l'ensemble des chemins possibles sur le graphe  $G(V(t), E(t), t)$ , commençant au minimum à  $t = d$  et se terminant avant  $t = d + \tau$ .

Nous réduisons le problème en utilisant une source unique comme point de départ (« single-source shortest path ») pour la recherche d'infecteurs : nous cherchons la longueur du (des) plus court(s) chemin(s) partant de l'infecteur  $s$  et pouvant rallier l'infecté  $t$ . Nous procédons par comptage en avant, en extrayant le sous-graphe égocentrique de  $s$  le jour  $d$ ,  $G_s(V_s(d), E_s(d), d)$ , jusqu'à ce que l'ensemble du composant connexe ayant pour source  $s$  soit parcouru. La matrice d'adjacence  $A_s^{(d)}$  associée à ce nouveau graphe sert de base pour l'ajout du sous-graphe connexe du jour  $d+1$  ayant pour source l'ensemble des nœuds  $V_s(d)$ .

Par itération jusqu'au réseau à  $d+\tau$ , nous obtenons un nouveau graphe  $G_s^{d \rightarrow d+\tau}(V_s^{d \rightarrow d+\tau}, E_s^{d \rightarrow d+\tau})$  dont la matrice d'adjacence  $A_s^{(d \rightarrow d+\tau)}$  permet de déduire l'ensemble des chemins partant de  $s$  entre  $d$  et  $d+\tau$ . Ce graphe est par ailleurs un sous-graphe du réseau agrégé naïvement et dont la matrice d'adjacence est  $A^* = \sum_{i=d}^{d+\tau} A^{(i)}$ .

Dans la suite de ce chapitre, l'identification de chaînes de transmission utilise ce calcul afin de déterminer s'il existe un chemin au moins entre un infecteur  $s$  (porteur au temps  $d$ ) et un infecté  $t$  (incident au temps  $d+\tau+1$ ). Les outils développés dans le cadre de notre analyse utilisent des matrices d'arêtes, dont la dimension est déterminée par le nombre d'arêtes dans le réseau, afin d'accélérer les calculs.

### 3.2.2.2 Calcul de la longueur de ces chemins

La matrice  $A_s^{(d \rightarrow d+\tau)}$  permet de calculer par itération la longueur du plus court chemin entre  $s$  et  $t$ , s'il existe. A partir du sous-graphe égocentrique ayant pour source  $s$ , nous parcourons à chaque itération  $k$  le voisinage  $k$ -hop de  $s$ , jusqu'à ce qu'il inclue  $t$  ou alors que l'ensemble du graphe ait été listé. Cette approche ne permet pas de lister le(s) intermédiaire(s) successif(s) allant de  $s$  à  $t$ , mais permet de compter le nombre minimum de sauts les séparant. Elle revient à sommer les puissances successives de la

matrice  $A_s^{(d \rightarrow d+\tau)}$  jusqu'à ce que l'élément  $a_{st}$  devienne non-nul (*i.e.* une partie de la matrice d'accessibilité), dans la limite de  $N$  puissances (nombre de nœuds).

### 3.3 Données de contacts détaillées et incidence de *S. aureus* en milieu hospitalier

L'étude que nous présentons dans cette section part du constat que les « contacts », au sens épidémiologique, sont souvent difficiles à définir et mesurés par des proxys. L'utilisation de capteurs de proximité, matériel peu coûteux et peu intrusif, est de plus en plus fréquente : les réseaux de contacts ainsi mesurés fournissent un proxy pour les interactions sociales au cours desquelles les participants se trouvent à moins d'une distance seuil définie par les investigateurs. Jusqu'alors, les études épidémiologiques émettent l'hypothèse que ce réseau a bel et bien un sens et peut servir de substrat pour analyser les comportements à risque et tester des mesures de contrôle. Toutefois, il n'existe pas encore de moyen de vérifier cette hypothèse. Nous nous proposons de la valider, en utilisant les données d'I-Bird pour développer une statistique permettant de tester la corrélation entre le réseau des contacts observés et le portage de *S. aureus*.

#### 3.3.1 Résumé de l'article en français

La collecte de données de contacts à haute résolution est de plus en plus fréquente pour étudier des phénomènes de dissémination de maladies.[10, 20, 79, 144] Ces réseaux sont utilisés afin de tester des stratégies d'immunisation ciblées ayant pour but de minimiser la taille finale de l'épidémie[130] : la modélisation de l'histoire naturelle d'un pathogène est faite selon un modèle compartimental, et la dissémination de ce pathogène se produit le long des arêtes du réseau. Toutefois, ces réseaux ne sont qu'un proxy pour les contacts menant à transmission. Au mieux, les contacts sont étudiés conjointement avec des données génotypiques, afin d'identifier *a posteriori* des chaînes de transmission plausibles.[58]

Dans cet article, nous développons une méthodologie tirant partie des doubles données de l'étude I-Bird, pour tester la corrélation entre le réseau des contacts enregistrés par des capteurs électroniques et le portage individuel hebdomadaire de *S. aureus*. Nous réalisons dans un premier temps des simulations microscopiques de dissémination, sur la base d'un modèle S-I-S informé par la collection de réseaux de contacts  $G(V(t), E(t), t)$ . Nous ajoutons à ces simulations un modèle d'observation permettant de censurer les données épidémiologiques afin de ressembler au mieux aux données hebdomadaires de I-Bird. Les simulations commencent à une date  $t$ , tirée au hasard parmi les dates où le

réseau a été enregistré, et continuent jusqu'à  $t+21$  jours. A l'état initial, un individu est choisi au hasard comme cas index pour un pathogène qui se propage de façon similaire à *S. aureus*. Nous définissons un épisode d'incidence comme l'acquisition du pathogène par un individu une semaine donnée, alors qu'il n'était pas colonisé la semaine précédente.

Les événements d'incidence détectés sont ensuite investigués afin de déterminer des infecteurs potentiels et leur distance sur le réseau.

L'hypothèse nulle du test que nous mettons en place est que le portage est indépendant des contacts. Cette hypothèse est simulée par une approche de Monte-Carlo, en permutant un grand nombre de fois le statut de tous les individus sauf du cas incident. Nous envisageons alors trois statistiques susceptibles de capturer la corrélation entre le réseau de contacts (ayant servi de substrat à la transmission) et l'acquisition du pathogène :

- (S1) le nombre de cas incidents pour qui un chemin existe jusqu'à un infecteur potentiel au moins,
- (S2) le nombre de cas incidents en contact direct avec leur infecteur(s) potentiel(s),
- (S3) la distribution de la distance topologique entre les cas incidents et leur(s) infecteur(s) potentiel(s).

Pour (S1) et (S2) nous utilisons un test du Chi-2, et pour (S3) le test des rangs signés de Wilcoxon pour données appariées. Nous comparons ces trois approches en évaluant leur puissance statistique pour un nombre croissant d'épisodes d'incidence.

Ces trois stratégies sont à mettre en regard avec les propriétés du réseau que nous étudions : la percolation s'y produit très fréquemment, et par conséquent la stratégie (S1) est incapable de détecter la corrélation entre le portage et les contacts. A l'inverse, les stratégies (S2) et (S3) se révèlent relativement performantes, atteignant rapidement des valeurs de puissances supérieures à 90%.

Nous appliquons finalement ces trois stratégies sur les données de portage originales de l'étude I-Bird. La définition d'un épisode d'incidence est affinée afin de tenir compte de possibles portages multiples et de la sensibilité imparfaite de la routine de détection (supposée entre 60% et 80%). Comme attendu, (S1) ne permet pas de mettre en évidence une différence significative entre l'hypothèse nulle et les données observées, car le réseau permet presque toujours de remonter jusqu'à un infecteur potentiel (97%

vs. 97%,  $P = 1$ ). Les stratégies (S2) et (S3) révèlent quant à elles que le réseau des contacts est bel et bien corrélé au portage individuel, démontrant que ce proxy électronique est adapté à l'étude des voies de dissémination de *S. aureus* en milieu hospitalier (et par extension à d'autres pathogènes aux modalités de transmission similaires).

**3.3.2 Article : Detailed Contact Data and the Dissemination of Staphylococcus aureus in Hospitals, Obadia T. et al., PLOS Computational Biology, 2015**



RESEARCH ARTICLE

# Detailed Contact Data and the Dissemination of *Staphylococcus aureus* in Hospitals

Thomas Obadia<sup>1,2\*</sup>, Romain Silhol<sup>3</sup>, Lulla Opatowski<sup>4,5,6</sup>, Laura Temime<sup>7</sup>, Judith Legrand<sup>8</sup>, Anne C. M. Thiébaud<sup>4,5,6</sup>, Jean-Louis Herrmann<sup>9,10</sup>, Éric Fleury<sup>11,12</sup>, Didier Guillemot<sup>4,5,6,13</sup>, Pierre-Yves Boëlle<sup>1,2,14</sup>\*, on behalf of the I-Bird Study Group<sup>¶</sup>

**1** Sorbonne Universités, UPMC Univ Paris 06, UMR\_S 1136, Institut Pierre Louis d'Epidémiologie et de Santé Publique, F-75013, Paris, France, **2** INSERM, UMR\_S 1136, Institut Pierre Louis d'Epidémiologie et de Santé Publique, F-75013, Paris, France, **3** Department of Infectious Disease Epidemiology, Imperial College London, London, United Kingdom, **4** Inserm UMR 1181 "Biostatistics, Biomathematics, Pharmacoepidemiology and Infectious Diseases" (B2PHI), F-75015, Paris, France, **5** Institut Pasteur, UMR 1181, B2PHI, F-75015, Paris, France, **6** Univ. Versailles St Quentin, UMR 1181, B2PHI, F-78180 Montigny-le Bretonneux, **7** Laboratoire MESuRS, Conservatoire National des Arts et Métiers, 75003, Paris, France, **8** Univ Paris-Sud, UMR 0320/UMR8120 Génétique Quantitative et Evolution—Le Moulon, F-91190, Gif-sur-Yvette, France, **9** INSERM U1173, UFR Simone Veil, Versailles-Saint-Quentin University, 78180, Saint-Quentin en Yvelines, France, **10** AP-HP, Hôpital Raymond Poincaré, Service de Microbiologie, F-92380, Garches, France, **11** ENS de Lyon, Université de Lyon, Laboratoire de l'Informatique du Parallélisme (UMR CNRS 5668—ENS de Lyon—UCB Lyon 1), IXXI Rhône Alpes Complex Systems Institute, Lyon, France, **12** Inria—Institut National de Recherche en Informatique et en Automatique, Montbonnet, France, **13** AP-HP, Raymond Poincaré Hospital, F-92380 Garches, France, **14** AP-HP, Hôpital Saint-Antoine, Département de Santé Publique, F-75571, Paris, France

\* These authors contributed equally to this work.

¶ Membership of the I-Bird Study Group is listed in the Acknowledgments.

\* [thomas.obadia@iplesp.upmc.fr](mailto:thomas.obadia@iplesp.upmc.fr) (TO); [pierre-yves.boelle@upmc.fr](mailto:pierre-yves.boelle@upmc.fr) (PYB)



## OPEN ACCESS

**Citation:** Obadia T, Silhol R, Opatowski L, Temime L, Legrand J, Thiébaud ACM, et al. (2015) Detailed Contact Data and the Dissemination of *Staphylococcus aureus* in Hospitals. PLoS Comput Biol 11(3): e1004170. doi:10.1371/journal.pcbi.1004170

**Editor:** Marcel Salathé, Pennsylvania State University, United States of America

**Received:** October 15, 2014

**Accepted:** February 3, 2015

**Published:** March 19, 2015

**Copyright:** © 2015 Obadia et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Data are available from the figshare database (<http://dx.doi.org/10.6084/m9.figshare.1290932>)

**Funding:** This study was supported by the European Commission under the Life Science Health Priority of the 6th Framework Program (MOSAR network contract LSHP-CT-2007-037941). Funding was also received from the French Government through the National Clinical Research Program and the Investissement d'Avenir program, Laboratoire d'Excellence "Integrative Biology of Emerging Infectious Diseases" (grant no. ANR-10-LABX-62-

## Abstract

Close proximity interactions (CPIs) measured by wireless electronic devices are increasingly used in epidemiological models. However, no evidence supports that electronically collected CPIs inform on the contacts leading to transmission. Here, we analyzed *Staphylococcus aureus* carriage and CPIs recorded simultaneously in a long-term care facility for 4 months in 329 patients and 261 healthcare workers to test this hypothesis. In the broad diversity of isolated *S. aureus* strains, 173 transmission events were observed between participants. The joint analysis of carriage and CPIs showed that CPI paths linking incident cases to other individuals carrying the same strain (i.e. possible infectors) had fewer intermediaries than predicted by chance ( $P < 0.001$ ), a feature that simulations showed to be the signature of transmission along CPIs. Additional analyses revealed a higher dissemination risk between patients via healthcare workers than via other patients. In conclusion, *S. aureus* transmission was consistent with contacts defined by electronically collected CPIs, illustrating their potential as a tool to control hospital-acquired infections and help direct surveillance.

IBEID). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

## Author Summary

Recent advances in communication technologies allow monitoring high-resolution contact networks. Close proximity interactions (CPIs) measured by wireless sensors are increasingly used to inform contact networks for the dissemination of pathogens in computational models, although empirical justification is lacking. Here, we conducted a longitudinal prospective study for four months in a hospital, including both patients and healthcare workers (HCWs). High-resolution CPIs were recorded continuously, and participants undertook weekly nasal swabs to detect *S. aureus* carriage. We set out to test whether the contact network measured by CPIs supported observed transmission episodes. A simulation study was first conducted to choose a test statistic for the association of CPI paths with transmission, showing that CPI path length from transmitter to incident case was the most powerful. Then, we selected patients presenting incident *S. aureus* colonization in the data. We showed that CPI paths existed to carriers of the same strain, with path lengths significantly shorter than between random pairs of participants, in agreement with the transmission hypothesis. In-hospital contact networks measured by CPIs inform on opportunities for pathogen transmission. These could be used in surveillance systems to help prevent the spread of nosocomial pathogens.

## Introduction

Chains of transmission in communicable diseases are often identified by *ad hoc* strategies, combining retrospective information on locations attended and pathogen genetics to identify time-consistent transmission paths.[1] In contrast with such undertakings, “digital epidemiology” propose to use new technologies to prospectively measure contacts and understand transmission[2,3]. Close-proximity interactions (CPIs) between persons recorded by wireless sensors[4] in real-life settings like schools or hospitals[5,6] have been used as indicators of contact in this respect. However, there is no evidence yet that such CPIs actually capture contacts explaining transmission.

To test this hypothesis, we designed a study where both *Staphylococcus aureus* carriage and CPIs were measured in a 200-bed long-term care facility with 5 wards. This setting has several advantages for our purpose: *S. aureus* is commonly found in healthcare facilities, colonizing patients and healthcare workers (HCWs); *S. aureus* carriage in the nares is usually prolonged as the nares are the most consistent area from which it can be isolated,[7] allowing its detection by repeated routine screenings; identical genetic and antibiotic-resistance profiles show that *S. aureus* strains spread among patients and HCWs[8,9]; long-term care facilities harbor a stable population, with patients staying for extended periods under the care of dedicated staff. The control of *S. aureus* transmission is also relevant for hospital hygiene, because its carriage increases the risk of healthcare-associated infections.[10]

In our study, *S. aureus* carriage was identified in patients and, importantly, in HCWs every week by repeated nasal swabbing. During the same period, all participants wore small wireless sensors that recorded their CPIs with each other in real time (every 30 s). To make the best use of this new data and account for the difference in temporal granularity, we first assessed the ability of several statistics to test the correlation of CPI records and *S. aureus* carriage. These characteristics are first presented based on the analysis of simulations where a pathogen spread according to the CPI network edges, then applied to the original data.

## Materials and Methods

### Ethics statement

Authorizations were obtained in accordance with French regulations regarding medical research and information processing. All French IRB-equivalent agencies accorded the i-Bird program official approval (CPP 08061; Afssaps 2008-A01284-51; CCTIRS 08.533; CNIL AT/YPA/SV/SN/GDP/AR091118 N°909036). Signed consent by patients and staff was not required according to the French Ethics Committee to which the project was submitted.

### Data collection

The I-Bird (individual-based investigation of resistance dissemination) study was conducted in a 200-bed long-term and rehabilitation hospital in northern France. The hospital is organized in 5 wards corresponding to medical specialties (geriatrics, neurology, nutrition, orthopedics, post-operative care). During the study period, 329 distinct patients stayed in the facility. Hospital staff (HCWs and other administrative personnel) numbered 261. In the text, all hospital staff is referred to as healthcare workers (HCWs). More details about the I-Bird investigation are provided in [S1 Text](#). When relevant, patients, nurses, nurses' aides and physicians were analyzed according to the ward in which they stayed or worked; night-shift staff, reeducation therapists and administrative personnel were excluded from these analyses as they were not assigned to a particular ward.

During the study period, all individuals (patients and HCWs) wore a small wireless sensor that recorded, every 30 s, the identity of other sensors that were in close proximity (typically < 1.5m, front-facing). The deployment of such sensors did not rely on any stationary infrastructure to record CPIs, as each sensor directly stored timestamped CPIs on its on-board flash memory. Further details on CPI collection and network reconstruction, along with descriptive characteristics of contact patterns, are available in [S2 Text](#). In the following, analyses are conducted on a dynamic CPI network aggregated at a daily scale by pair of individuals (*ie* network edges). We defined an individual's  $k$ -hop neighborhood as all other individuals in the network who were found within  $k$  steps from him. For example, the 1-hop neighborhood of an individual contained all his direct neighbors, while his 2-hop neighborhood contained both his direct neighbors and these neighbors' neighbors.

### Defining incident *S. aureus* colonization episodes and CPI-supported transmission paths

All participants underwent weekly nasal swabs to monitor *S. aureus* carriage. Upon detection of *S. aureus* colonization, the isolated strains were *spa*-typed [[11,12](#)] and their resistance profiles to 20 antibiotics were determined (see also [S1 Text](#) for detailed protocol). The screening procedure had an expected sensitivity of 61.5% and specificity of 98.8%, [[13](#)] although higher sensitivity value has been reported (~80%). [[14](#)] The anterior nares were preferred to other body areas because they harbor the most stable *S. aureus* colonization and also reflect on overall body carriage. [[7,15](#)] Furthermore, eradication of nasal carriage is also associated with eradication of skin carriage. [[16,17](#)] *S. aureus* strains were considered identical when they had the same *spa* type and antibiotic-resistance profile, in accordance with studies comparing *spa* typing to other molecular techniques. [[11,12,18,19](#)] Transmission events were identified by the isolation of a new *S. aureus* strain from a patient's swabs, defining "incident colonization episodes". Because HCWs may be transiently colonized, [[8,20,21](#)] which would mostly be missed with weekly swabs, we only considered incidence in patients. To account for imperfect *S. aureus* detection in case of multiple carriage, we also required that the new *S. aureus* strain had not been

detected in the patient's previous 2 swabs had he been colonized with another strain in the preceding week.

Each incident colonization episode was investigated to identify time-consistent CPI paths linking the incident case to a previous carrier of the same strain. Recent CPI paths were favored over others by applying the following algorithm:

- All individuals carrying the same strain in the three preceding weeks were defined as “candidate transmitters”, regardless of their CPI connections.
- CPI paths to all candidate transmitters were looked for. In case of existence, the candidate transmitter became a “CPI-supported candidate transmitter”. In this case, the CPI path length (in hops) was computed.
- We sorted all CPI-supported candidate transmitters according to distance in time, then distance in hops. The first CPI-supported candidate transmitter in this list was the “CPI supported transmitter”. In other words, it was the least remote in number of hops among all candidate transmitters arising the least remote in time. In case of ex aequo, one of the candidate transmitters was chosen at random if required.

We investigated 3 weeks before incidence as it allowed finding a CPI-supported transmitter for all incident episodes but 4, which were not CPI-supported by exploring further back in time.

## Statistical analysis

**Testing for CPI supported transmission: test statistics & simulations.** We identified three observable quantities that would provide evidence for the correlation between CPIs and observed transmission: S1—The proportion of incident colonization episodes with least one CPI-supported transmitter; S2—The proportion of CPI-supported transmitters in direct CPI with an incident case; and S3—the length of the shortest CPI-supported transmission path (a good proxy to the actual transmission path[22]). Each of these characteristics can be used to build a test, where, classically, observations would be compared to the expected values under the null hypothesis of independence between CPIs and transmission. These expected values can be computed by a Monte Carlo approach: carriage information was first randomly permuted between participants. To keep autocorrelation between successive swabs in the same individuals, we permuted carriage information over the 3 preceding weeks simultaneously. As *S. aureus* prevalence was similar between patients and HCWs (Table 1), we did not take occupation into account for permutations. For each incident colonization episode, 100 replicates of permuted carriage statuses were generated to simulate the distribution of statistics of interest for investigated strategies.

For S1 and S2, we compared the observed percentages of CPI supported paths to that expected after random permutations of carriage (e.g. in S1, the observed percentage of CPI-supported episodes was compared to the average proportion of CPI-supported episodes among all permutations). For S3, the shortest CPI path length was averaged across all replicates for each incident colonization episodes, thus providing the expected distribution under the null. The observed CPI path length distribution was then compared to that expected under the null using the Wilcoxon signed rank paired test.

To first study the characteristics of the three approaches and choose the most powerful, we used a simulation study based on a Susceptible—Colonized—Susceptible transmission model on the CPI network. Stochastic simulations were performed to mimic observations of incident colonization episodes in our study. First, the dynamic CPI network between all participants in

**Table 1. Characteristics of participants, close proximity interactions and *S. aureus* carriage.**

	Patients	HCW
<b>Total number</b>	329	261
<b>Sex (% male)</b>	43.5%	42.5%
<b>Age (median [range])</b>	58.5 [24.5–102.8]	41.3 [18.7–61.3]
<b>CPI number per d</b>	12 (± 6.2)	15 (± 7.2)
<b>With patients</b>	6 (± 4.5)	9 (± 5.7)
<b>With HCWs</b>	6 (± 2.5)	6 (± 3.1)
<b>CPI cumulative duration (h/d)</b>	12.2 (± 11.3)	3.7 (± 2.4)
<b>With patients</b>	11.1 (± 11)	1.7 (± 1.3)
<b>With HCWs</b>	1.1 (± 1.6)	2 (± 1.7)
<b>1-hop neighborhood</b>		
<b>Size</b>	76 (± 48)	NA
<b>Time to 50%</b>	5 days (± 8)	NA
<b>Time to 75%</b>	14 days (± 14)	NA
<b><i>S. aureus</i> carriage prevalence</b>	38% [35.3–40.7%]	36.3% [32.3–40.4%]
<b>1-month cumulative <i>S. aureus</i> incidence</b>	33% [25–41%]	NA

Values are mean ± SD or percent [95% C.I.], unless stated otherwise. NA, not applicable.

doi:10.1371/journal.pcbi.1004170.t001

a random 3-week long period was selected. All individuals in this network were assumed initially noncolonized (i.e. susceptible), except one randomly chosen to be the initial carrier in the first week. For each day  $d$  in the following 3 weeks, a noncolonized individual (say  $i$ ) could become colonized with probability

$$P_i(d) = 1 - [(1 - P_{PA})^{n_{PA}(i,d-1)} * (1 - P_{HCW})^{n_{HCW}(i,d-1)}]$$

where  $n_{PA}(i, d-1)$  is the number of carrier-patient neighbors on day  $d-1$  and  $n_{HCW}(i, d-1)$  the number of carrier-HCW neighbors of  $i$ ,  $P_{PA}$  the probability of transmission per contact with a carrier patient and  $P_{HCW}$  with a HCW. Colonized individuals cleared colonization at a constant rate  $q_{PA} = 0.1 \text{ days}^{-1}$  for patients and a  $g_{HCW} = 0.45 \text{ days}^{-1}$  for HCWs in agreement with other studies.[23,24] The observation model closely imitated those of our investigation: from the wholly simulated transmission chain, we only selected data determining carriage once a week in each participant. As in practice, the status of participants in the same ward was determined the same day. Finally, an incident colonization episode was selected from new carriers in week 3 of the simulation, as in the original data. Simulations were run to produce observations of a variable number of incident cases. The power of each statistical approach, S1 to S3, was determined as a function of the number of incident episodes.

**Survival analysis of *S. aureus* colonization in patients.** The time to colonization in patients was studied by survival analysis. Time was counted from admission to first colonization episode and censored at discharge otherwise. A Cox proportional hazard regression model was used to assess the risk of colonization with respect to network-related covariates, including the numbers of weekly CPIs (*ie* an individual weekly degree) and their cumulative durations, which were included as time-dependent.

**Risk of transmission through patients or HCWs.** The relative risk (RR) of transmission via a HCW *versus* a patient was estimated by the ratio of the percentage of intermediary HCWs



in CPI paths leading to transmission to that not leading to transmission. For each candidate transmitter at 2 hops from an incident case, we determined how many 2-hop neighbors were in the CPI network, and whether they were incident cases.

All statistical analyses were conducted with R v3.0.1.

## Results

CPIs were recorded among 590 individuals (329 patients and 261 HCWs) during the 4-month period (Table 1), and yielded 85,025 daily CPIs. Each day, a CPI network was defined with study participants as nodes and CPIs as edges (Fig. 1). The collection of daily CPI networks defined the “dynamic CPI network”. While the numbers of CPIs were within the same range for patients and HCWs, the daily-cumulative durations of the CPIs were much longer for patients than HCWs, respectively: 12.2 ( $\pm$  11.3) h (mean  $\pm$  SD) vs. 3.7 ( $\pm$  2.4) h (Fig. 2). Further description of the dynamic CPI network is provided in S2 Text.

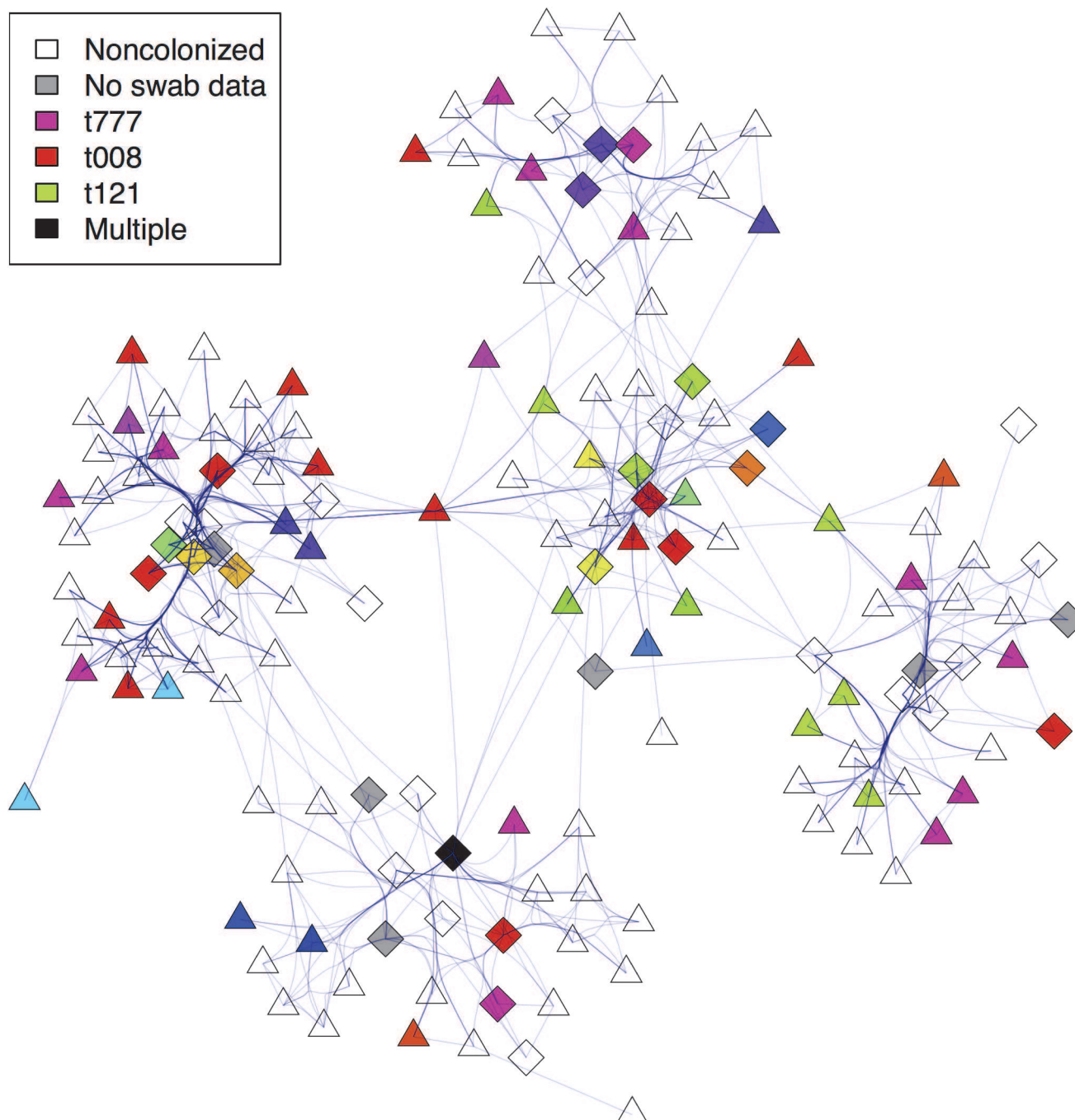
During the same time, 4,175 swabs were collected: 37.2% of them (1550 swabs from 363 participants) were positive for *S. aureus* carriage. In all, 148 different *spa* types were isolated during the study (Fig. 1, and Fig. 1 in S1 Text for incidence per week and *spa* types). Notably, 114 strains were isolated more than once, each in three participants on average, suggesting that transmission had occurred among these individuals.

## A simulation study of tests for the correlation of CPI-supported paths and carriage

We assessed the power of all proposed strategies with increasing numbers of incident colonization episodes (10, 20, 30, 50, 100 and 153). For each of these amounts, 500 replicates of 100 permutations were performed. Each strategy was performed on every replicate. Table 2 shows the power of the tests to reject the null hypothesis of independence between CPIs and transmission. Strategy S1, based on the existence of CPI supported transmitters, yielded very poor results. Indeed, in almost all situations, a CPI-supported transmitter existed in the original data as well as in the permuted data, so that no difference from random was seen in this characteristic. The percentage of incident cases in direct CPI with CPI-supported transmitters and the shortest CPI-supported path length yielded more useful procedures. The length of the shortest CPI path from transmitter to incident case (S3) was slightly more powerful than the percentage of transmitters in direct CPI with the incident case (S2), although both approaches had large power for rejecting the null for samples of size 153.

## Carriage of *S. aureus*

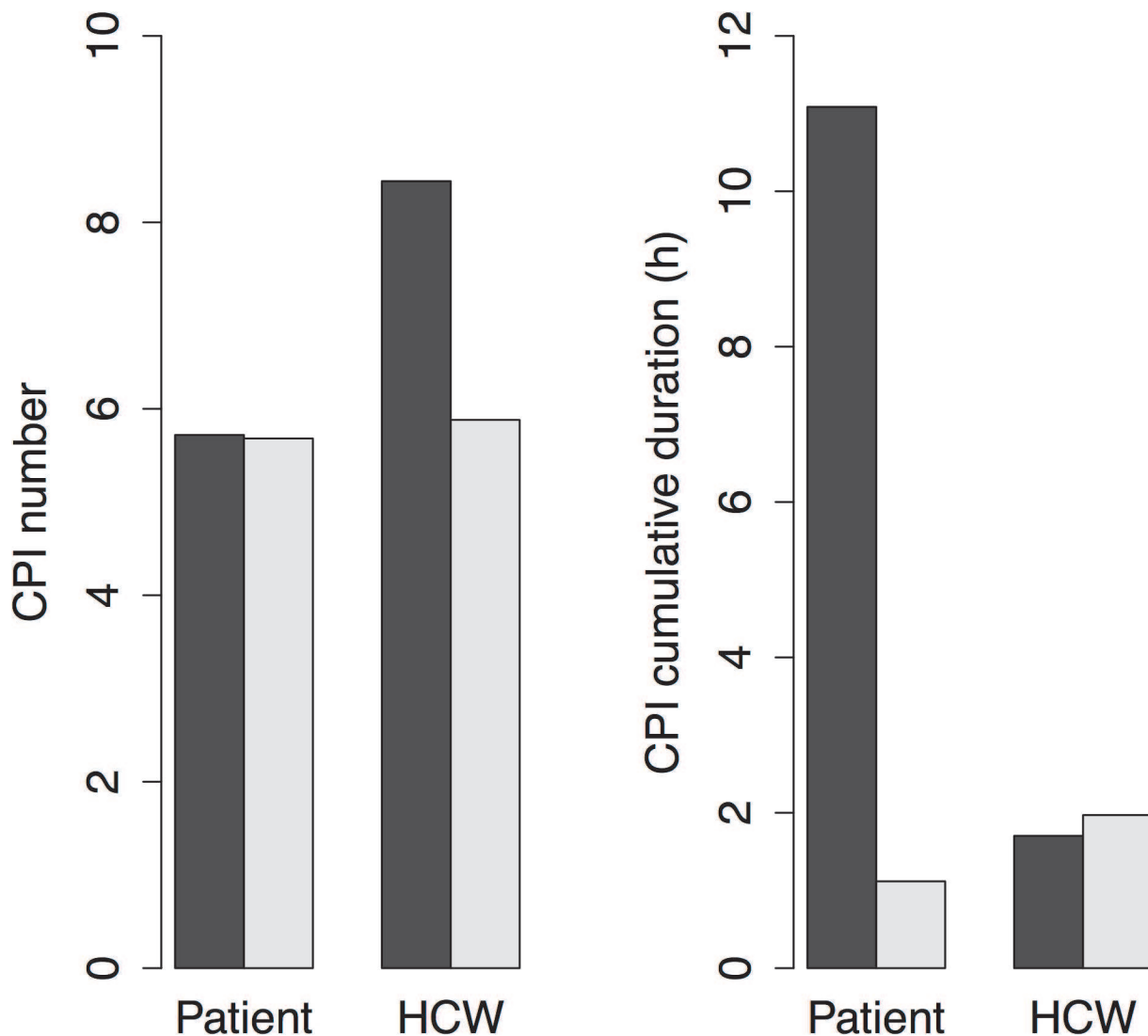
The time to first *S. aureus* colonization was analyzed for 201 patients who were not colonized at admission: 73 experienced incident colonization. The cumulative incidence of *S. aureus* colonization was 33% (95% confidence interval (C.I.) [25–41%]) 1 month after admission, with almost equal incidence of methicillin-resistant *S. aureus* (MRSA) (23.2% (95% C.I. [15.1–30.6%])) and methicillin-sensitive *S. aureus* (MSSA) (16.5% (95% C.I. [9.3–23.2%])). The risk of colonization did not change with the number of distinct direct neighbors during the preceding week (*ie* weekly degree; hazard ratio (HR) = 1.05 (95% C.I. [0.95–1.21]) for a 5-neighbor increase,  $P$  = 0.4), using either the raw number of CPIs (*ie* the sum of daily degrees; HR = 1 (95% C.I. [0.95–1.10]) for a 5-CPI increase,  $P$  = 0.4) or the cumulative duration of CPIs (*ie* the weight of network edges; HR = 1 (95% C.I. [0.99–1.01]) for a 1-h increase,  $P$  = 0.6). The same conclusions were drawn for MSSA or MRSA colonizations.



**Fig 1. CPI-based network and *S. aureus* carriage in the hospital.** The network shown corresponds to interactions occurring during 1 d. Patients are shown as triangles and HCWs as diamonds. Color-coding corresponds to the *spa* type of the last known colonization status (i.e. the preceding week). Because of the large *S. aureus* *spa* types diversity, only the 3 most common are reported in the legend. We used Fruchterman-Reingold force-directed algorithm for the layout (individuals in the network are closer together as the density of links among them increases). Force-directed edge bundling was used to accommodate their high density.

doi:10.1371/journal.pcbi.1004170.g001

Overall, 237 incident-colonization episodes were documented in 111 patients (144 MRSA, 93 MSSA). For each incident episode, we identified “candidate transmitters”, i.e. people who had carried the same strain at any time in the preceding three weeks. Among the 237 incident-colonization episodes, 173 (73%) had 307 candidate transmitters, with no difference between MRSA and



**Fig 2. Mean daily numbers and durations of patients' and HCWs' CPIs.** Bars illustrate CPIs with patients (black) and HCWs (white).

doi:10.1371/journal.pcbi.1004170.g002

MSSA episodes (76% (110/144) vs. 68% (63/93),  $P = 0.16$ ). Episodes without a candidate transmitter did not occur earlier post-admission than others (8.1 vs. 8.3 weeks,  $P = 0.8$ ), or preferentially in some wards ( $P = 0.13$ ). Twenty of the 173 episodes with candidate transmitters were discarded as the incident case had a missing CPI record due to sensor failure in the preceding weeks.

**Table 2. Statistical power computed for the three proposed test statistics.**

	<i>n</i> = 10	<i>n</i> = 20	<i>n</i> = 30	Power (%) <i>n</i> = 50	<i>n</i> = 100	<i>n</i> = 153
<b>Test statistics</b>						
CPI-supported paths (S1)	0	0	0	0	0	0
CPI-supported transmitter in direct contact (S2)	0.67	0.93	0.98	0.99	1	1
CPI-supported transmitter path length (S3)	0.75	0.96	0.99	1	1	1

Power was determined for an increasing set of incident colonization episodes, with 500 replicates each time.

doi:10.1371/journal.pcbi.1004170.t002



## Testing for CPI support

We investigated all 153 incident colonization episodes detected from longitudinal swab data. As expected from the simulations, a CPI path existed between the candidate transmitter and the incident case in almost all instances (97%, Table 3). The characteristics of the shortest CPI paths lengths from candidate transmitter to incident case were in favor of transmission along CPIs, with shorter paths in the original data than after random permutations (Fig. 3, Strategy S3:  $P < 0.001$ ). This was therefore the sign that *S. aureus* transmissions detected from the I-Bird swabs were driven by CPIs.

A direct CPI contact existed between the candidate transmitter and incident case (i.e. a CPI path of length 1) in 48% of the cases vs. only 30% expected by chance (Strategy S2:  $P < 0.0001$ ). A CPI path of length 2 was observed in 38% (vs. 53%) of the episodes and of length larger than 2 in the rest of the cases.

In most cases (64%), a CPI supported transmitter was found in the preceding week. The remaining were found in the preceding two (23%) or three weeks (13%).

## Increased transmission via HCWs

We next investigated transmission differences according to occupation, focusing on dissemination events for which the path from candidate transmitter to incident case had exactly one (noncolonized) intermediary. For these 2-hop CPI paths, *S. aureus* spread was more frequently observed when the intermediary was a HCW than a patient (2.1% vs. 1.8%,  $P = 0.0004$ ). The relative risk (RR) of transmission by a HCW was therefore 1.2 (95% C.I. [1.1–1.3]). This increased risk was more pronounced when the initial carrier was a patient (1.9% vs. 1.5%,  $P < 0.0001$ ; RR = 1.3 (95% C.I. [1.1–1.4])) rather than an HCW (3.2% vs. 3.0%,  $P = 0.48$ ; RR = 1.1 (95% C.I. [0.9–1.2])).

## Sensitivity analyses

All analyses were repeated using a thinned dynamic CPI network obtained by excluding short individual interactions lasting  $< 5$  min prior to daily aggregation. The thinned CPI network was still dense, including 8.1% of all potential interactions. As expected, this decreased density increased the number of intermediaries between 2 persons (mean =  $11 \pm \text{SD} = 6$ ) compared to the full network and decreased the number (mean =  $4 \pm \text{SD} = 1$ ) and duration of daily CPIs. The distribution of within/outside-ward CPIs was almost the same as before, with 75% of CPIs

**Table 3. Incident-colonization episodes, candidate transmitters and CPI support.**

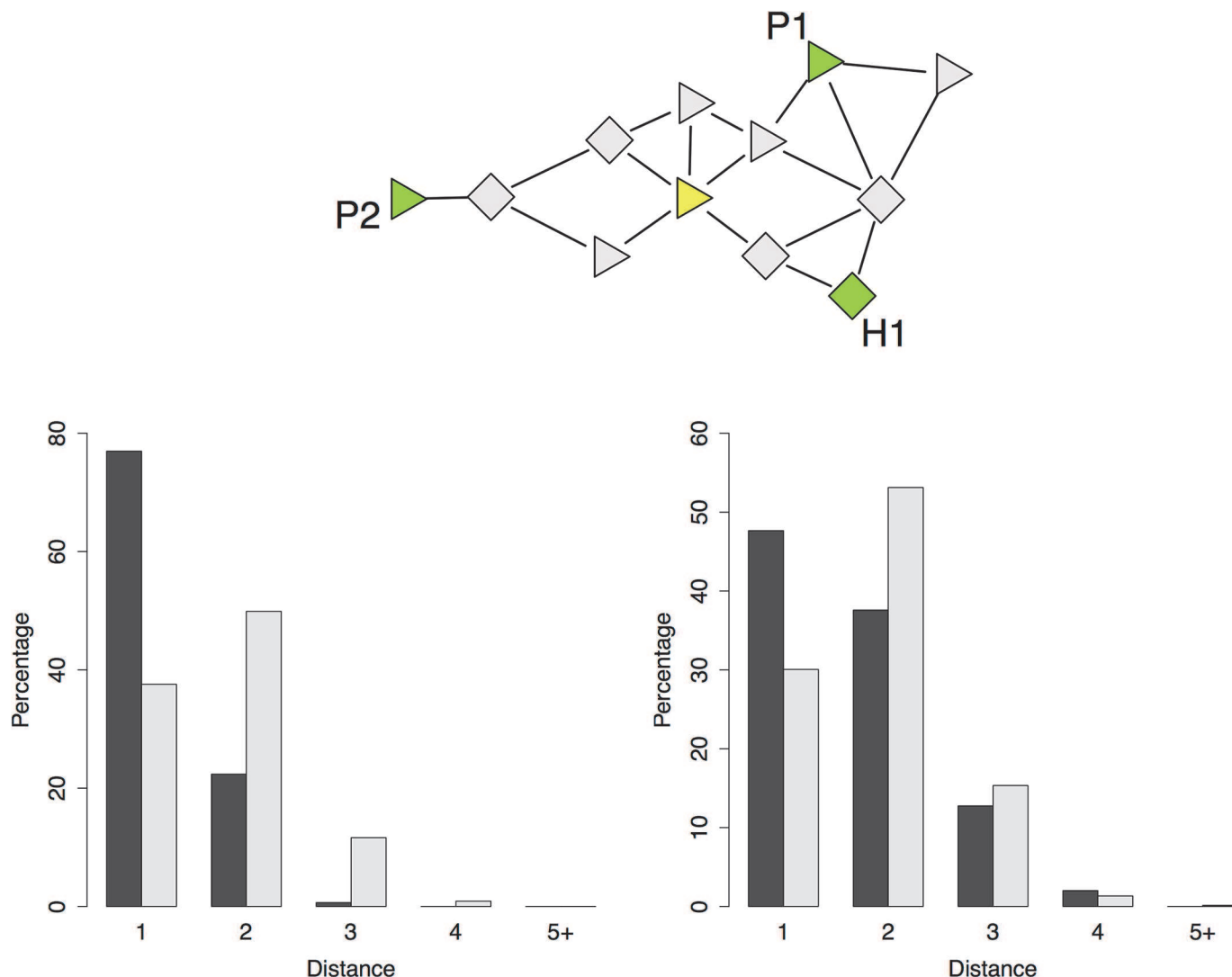
Incident-colonization episode	<i>S. aureus</i>	MRSA	MSSA
All, n	237	144	93
With $\geq 1$ candidate transmitter(s), n	173	110	63
With available CPI network*, n	153	100	53
CPI-supported†, n	149 (97%)	99 (99%)	50 (94%)

Incident-colonization episodes were investigated when the incident strain had been isolated from at least one another participant (candidate transmitter). When a CPI path existed between a candidate transmitter and the incident case, the episode was considered as CPI-supported.

\* Episodes where CPI were not recorded for the incident case were discarded

† Among those with available CPI network

doi:10.1371/journal.pcbi.1004170.t003



**Fig 3. CPI supported transmitters.** (Top) CPI-supported transmitters were selected among carriers of the same strain (green) as the incident case (yellow) who were the closest in the CPI network. Here, P1 and H1 are two CPI-supported transmitters in the incident case's 2-hop neighborhood, but not P2 who is further away (3-hop neighborhood). Patients are shown as triangles and HCWs as diamonds. (Bottom left) Comparison of the distance distribution between CPI-supported transmitters and incident cases in the data (black) and with random permutations of carriage data (white) in the simulation study. (Bottom right) Comparison of the distance between CPI-supported transmitters and incident cases in the original data.

doi:10.1371/journal.pcbi.1004170.g003

occurring within the ward. Again, no risk factor that could increase the risk of colonization was identified.

The CPI support of transmission was even larger than before, with shorter path lengths in the original network than expected by chance ( $P < 0.0001$ ). In the thinned network, 64% of candidate transmitters were in the incident case's 2-hop neighborhood (i.e. direct or with one intermediary), compared to 46% with the permuted carriage data.

Keeping all candidate transmitters with a path to the incident case, rather than only the closest ones did not change the conclusions. Excluding repeated incident episodes of the same strain in a given patient, leading to consider 129 transmission episodes only, did not affect the conclusions drawn regarding CPI support. Finally, keeping only episodes that were CPI-

supported in the first week before incidence led to the same conclusion (96 CPI-supported episodes, CPI path length significantly shorter ( $P = 0.01$ )).

To account for imperfect sensitivity and possible false negatives in swabs, we discarded all incident episodes where the carriage had been positive/negative/positive with the same strain at both ends, as those may be false negatives. This led to retain 129 incident colonization episodes with candidate transmitters and CPI records, among which 126 were CPI supported (ie 98%). The shortest CPI path length was significantly shorter than chance predicted ( $P < 0.0001$ ).

## Discussion

The contribution of the contact network between patients and HCWs in explaining hospital-associated infections is widely accepted[25,26], although it has never been tested empirically. Here, using electronic wireless devices to record close proximity interactions among persons in a hospital, we find evidence that these interactions are indeed informative for *S. aureus* transmission. To date, such high-resolution contact networks were used to structure contacts in computerized models and study how characteristics of individuals[27,28] and network topology[2,29] could influence the course of outbreaks. However, these findings relied on the underlying hypothesis that CPI paths actually captured dissemination paths. Our results provide evidence that CPIs recorded by electronic sensors are indeed relevant to explain transmission. This validates using such networks in future epidemiological studies, and should provide a powerful tool to better characterize risk and plan control measures for pathogens transmission in specific settings.

Recording interactions among individuals is increasingly easy using remote sensors[30]. In the study facility, wearing sensors was well accepted by patients and HCWs. Because the recording was limited (typically  $< 1.5$  m), these signals may be a good proxy for real-life contacts between people. Although contact surveys were shown to document contacts hardly overlapping with those recorded by electronic sensors,[31] direct observations by dedicated investigators provided more congruent data.[32] Individual surveys tend to omit contacts of short duration,[31] and therefore electronic sensors might capture more complete data regarding contact patterns. In our study, although most CPIs occurred within the wards (leading to clustered communities, as seen in Fig. 1), the CPI network was rather dense, covering up to 20% of all possible interactions among participants. The shortest path between any two individuals had few intermediaries, a typical “small-world” feature[33]. HCWs spent approximately 20% of their work shifts in direct contact with patients (1h50 out of 8h), in the same range as that reported in an emergency unit ( $\sim 30\%$ )[34] and the cumulated CPIs duration in a HCW compared with that reported for another ICU ( $\sim 2.2$  h)[32]. However, in sharp contrast with an earlier study conducted in a pediatric setting[30], where almost no contacts existed between patients, CPIs between patients herein were frequent and prolonged, as expected in a long-term care and rehabilitation facility where patients can initiate social interaction with others more easily than in acute-care hospital. Finally, ward organization was important in structuring contacts between patients and HCWs. This contact-clustering suggests that some interaction patterns between patients and HCWs could be rather constant from one hospital to the next, e.g., the numbers and durations of interactions, but that the full contact network may depend on type of care and ward organization. Other features of interest were the quick encounter of most of direct contacts during the first week after admission and that an incoming patient’s 3-hop neighborhood quickly encompassed almost all individuals in the hospital. This small-world feature may profoundly affect the potential spread of pathogens, increasing the size and speed of outbreaks[33,35].

*S. aureus* carriage was common among patients and HCWs, with a significant percentage of patients already colonized at admission (33.7% for *S. aureus*; 17% for MSSA; 18.9% for MRSA). Approximately one-third of noncolonized, newly admitted patients became colonized with *S. aureus* within the first month post-admission, as frequently observed in long-term care facilities[36,37]. The cumulative MRSA and MSSA incidence rate were similar after 1 month, as were their mean carriage prevalence, suggesting little, if any, difference in transmission between resistant and non-resistant strains. Patients admitted to long-term care facilities come from other hospitals and prevalence of carriage at admission is large. We analyzed incident carriage episodes only to focus on transmission occurring within the long-term care facility. The numbers and durations of contacts, although pointing towards increased risk for participants with more and longer contacts, were not by themselves strongly associated with *S. aureus* colonization: additional information on whether the contacts were carriers is likely to be required in this respect. Finally, because only weekly nasal swabs were conducted, neither hand carriage nor transient colonization episodes (<1 week) were considered, although hand carriage by HCWs has been described[8]. In our analysis, this may lead to imperfect observation of the carriage status, as participants may have cleared colonization between successive swabs.

In our study, *S. aureus* carriage was determined by nasal swabs. As previously stated, skin colonization was not detected. Yet, previous studies have shown that while *S. aureus* can be isolated from many anatomic location, nasal swabs were consistent with carriage isolated from other area of the body in 82% of the case[38]. In HCWs, where transient colonization is more likely to occur, this might lead to underestimate prevalence and therefore overestimate incidence, as an (unobserved) skin carriage could lead to a longer, more stable, colonization of the nares. For this reason, we chose not to include incident colonization episodes occurring in HCWs. Although the main route of transmission for *S. aureus* remains physical contact with carrier individuals, transmission through the environment is also possible, for example in the form of fomites.[26] Our procedure cannot distinguish between routes of transmission leading to short CPI paths, for example, if contact with fomites occurred only when people were at CPI range of a known carrier. Yet, our results suggest that CPIs, as defined in our study setting, correlated with *S. aureus* transmission routes and are therefore a good proxy[35] for interactions leading to *S. aureus* dissemination.

The choice of a test statistic to test for transmission along CPIs required giving some consideration to the setup of this study. First, the density of contacts between participants was large, as in occupational networks measured in health care structures.[24,30] This leads to percolation,[39] with typical short distance between participants. Therefore, a CPI path between a *S. aureus* carrier and an incident case was not specific enough of transmission: it was the rule rather than the exception, explaining why strategy S1 could not evidence association. Second, while CPIs could be recorded continuously, it is obvious that *S. aureus* carriage cannot be determined as frequently. Some participants could therefore clear carriage between two successive swabs, hiding their role in transmission. In this case, direct CPI would not be seen in the recorded CPIs. However, a short CPI path may be found to a more distant carrier through non-colonized intermediaries and would still be supportive of transmission. The shortest CPI path between CPI-supported transmitters and incident cases allowed to account for carriage gaps in observed transmission chains and was more informative than mere existence of a connecting path. The power comparison between strategies S1, S2 and S3 showed that this was indeed the case. It also evidenced that the proposed tests actually discriminated between transmission along the CPI paths and random transmission with no relationship to the CPI network.

In contrast to weekly swabbing, wireless sensors recorded interactions permanently and made it unlikely that the network of interactions was imperfectly observed. *S. aureus*

transmission is thought to occur mainly through physical contacts and these should be present in the CPI recordings. However, the CPI network may capture additional interactions that are unlikely to lead to transmission. To focus on interactions that were the most likely to lead to transmission, for example nursing care, we discarded all CPIs lasting less than 5 min. In this thinned network, we found again evidence that paths defined by CPIs supported transmission. Finally, the increased likelihood of *S. aureus* spread through a—seemingly noncolonized—HCW intermediary was in good agreement with their importance in transmission and the occurrence of transient colonization among them.

The hypothesis that CPIs are a good proxy for contacts leading to transmission of *S. aureus* is highly plausible *a priori*. Indeed, the main route of transmission for *S. aureus* is physical contact; those led to CPI records as sensors recorded all physical proximity (<1.5m). The mechanism of transmission was therefore captured by CPIs. With the evidence we provide on the correlation between CPI paths and transmission events, this strengthens the interest of this proxy measure as a cheap, feasible and informative method for studying *S. aureus* transmission.

## Conclusion

The joint analysis of *S. aureus* carriage and CPI data collected during 4 months provided evidence that CPIs capture contacts associated with transmission. This supports using CPI information to improve the realism of transmission models. This also suggests that a more systematic in-depth study of CPI networks could provide new directions for controlling *S. aureus* transmission in hospitals.

## Supporting Information

**S1 Text. I-Bird investigation methods.**

(DOCX)

**S2 Text. Description of the dynamic CPI network.**

(DOCX)

## Acknowledgments

We thank Janet Jacobson for advice and editorial assistance.

**I-Bird (Individual-Based Investigation of Resistance Dissemination) study group members**

Anne Sophie Alvarez (AP-HP, Paris, France), Audrey Baraffe (AP-HP, Paris, France), Mariano Beiró (Universidad de Buenos Aires, Buenos Aires, Argentina), Inga Bertucci (AP-HP, Paris, France), Pierre-Yves Boëlle (Univ. Pierre et Marie Curie, Paris, France), Camille Cyncynatus (AbAg, Chilly-Mazarin, France), Florence Dannet (AP-HP, Paris, France), Marie Laure Delaby (AP-HP, Paris, France), Pierre Denys (AP-HP, Paris, France), Matthieu Domenech de Cellès (Univ. Pierre et Marie Curie, Paris, France), Eric Fleury (ENS Lyon, Lyon, France), Antoine Fraboulet (Insa, Lyon, France), Jean-Louis Gaillard (AbAg, Chilly-Mazarin, France), Didier Guillemot (Univ. Versailles Saint-Quentin, Versailles, France), Jean-Louis Herrmann (Univ. Versailles Saint-Quentin, Versailles, France), Boris Labrador (AP-HP, Paris, France), Jennifer Lasley (Inserm, Paris, France), Christine Lawrence (AP-HP, Paris, France), Judith Legrand (Univ. Paris Sud, Orsay, France), Odile Le Minor (Institut Pasteur, Paris, France), Caroline Ligier (Institut Pasteur, Paris, France), Lucie Martinet (Inria, Lyon, France), Karine Mignon (AbAg, Chilly-Mazarin, France), Lulla Opatowski (Univ. Versailles Saint-Quentin, Versailles, France), Catherine Sacleux (AP-HP, Paris, France), Jérôme Salomon (Cnam, Paris, France), Laura Temime (Cnam, Paris, France), Thomas Obadia (Univ. Pierre et Marie Curie, Paris,

France), Marie Perard (AP-HP, Paris, France), Laure Petit (Institut Pasteur, Paris, France) Laetitia Remy (AP-HP, Paris, France), Anne Thiebaut (Inserm, Paris, France), Damien Thomas (AbAg, Chilly-Mazarin, France), Philippe Tronchet (AP-HP, Paris, France), Isabelle Villain (AP-HP, Paris, France).

## Author Contributions

Conceived and designed the experiments: DG EF PYB JLH. Analyzed the data: TO PYB RS. Wrote the paper: TO PYB RS JL LO LT ACMT JLH EF DG.

## References

1. Gardy JL, Johnston JC, Sui SJH, Cook VJ, Shah L, Brodtkin E, et al. Whole-Genome Sequencing and Social-Network Analysis of a Tuberculosis Outbreak. *N Engl J Med*. 2011; 364: 730–739. doi: [10.1056/NEJMoa1003176](https://doi.org/10.1056/NEJMoa1003176) PMID: [21345102](https://pubmed.ncbi.nlm.nih.gov/21345102/)
2. Salathé M, Kazandjieva M, Lee JW, Levis P, Feldman MW, Jones JH. A High-Resolution Human Contact Network for Infectious Disease Transmission. *Proc Natl Acad Sci*. 2010; 107: 22020–22025. doi: [10.1073/pnas.1009094108](https://doi.org/10.1073/pnas.1009094108) PMID: [21149721](https://pubmed.ncbi.nlm.nih.gov/21149721/)
3. Salathé M, Bengtsson L, Bodnar TJ, Brewer DD, Brownstein JS, Buckee C, et al. Digital Epidemiology. *PLoS Comput Biol*. 2012; 8: e1002616. doi: [10.1371/journal.pcbi.1002616](https://doi.org/10.1371/journal.pcbi.1002616) PMID: [22844241](https://pubmed.ncbi.nlm.nih.gov/22844241/)
4. Cattuto C, Van den Broeck W, Barrat A, Colizza V, Pinton J-F, Vespignani A. Dynamics of Person-to-Person Interactions from Distributed RFID Sensor Networks. *PLoS ONE*. 2010; 5: e11596. doi: [10.1371/journal.pone.0011596](https://doi.org/10.1371/journal.pone.0011596) PMID: [20657651](https://pubmed.ncbi.nlm.nih.gov/20657651/)
5. Hashemian M, Stanley K, Osgood N. Leveraging H1N1 infection transmission modeling with proximity sensor microdata. *BMC Med Inform Decis Mak*. 2012; 12: 35. doi: [10.1186/1472-6947-12-35](https://doi.org/10.1186/1472-6947-12-35) PMID: [22551391](https://pubmed.ncbi.nlm.nih.gov/22551391/)
6. Salathé M, Jones JH. Dynamics and Control of Diseases in Networks with Community Structure. *PLoS Comput Biol*. 2010; 6: e1000736. doi: [10.1371/journal.pcbi.1000736](https://doi.org/10.1371/journal.pcbi.1000736) PMID: [20386735](https://pubmed.ncbi.nlm.nih.gov/20386735/)
7. Williams REO. Healthy Carriage of Staphylococcus Aureus: Its Prevalence and Importance. *Bacteriol Rev*. 1963; 27: 56–71. PMID: [14000926](https://pubmed.ncbi.nlm.nih.gov/14000926/)
8. Blok HEM, Troelstra A, Kamp-Hopmans TEM, Gigengack-Baars ACM, Vandenbroucke-Grauls CMJE, Weersink AJL, et al. Role of Healthcare Workers in Outbreaks of Methicillin-Resistant Staphylococcus aureus: A 10-year Evaluation From a Dutch University Hospital. *Infect Control Hosp Epidemiol*. 2003; 24: 679–685. PMID: [14510251](https://pubmed.ncbi.nlm.nih.gov/14510251/)
9. Köser CU, Holden MTG, Ellington MJ, Cartwright EJP, Brown NM, Ogilvy-Stuart AL, et al. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N Engl J Med*. 2012; 366: 2267–2275. doi: [10.1056/NEJMoa1109910](https://doi.org/10.1056/NEJMoa1109910) PMID: [22693998](https://pubmed.ncbi.nlm.nih.gov/22693998/)
10. Kluytmans J, van Belkum A, Verbrugh H. Nasal carriage of Staphylococcus aureus: epidemiology, underlying mechanisms, and associated risks. *Clin Microbiol Rev*. 1997; 10: 505–520. PMID: [9227864](https://pubmed.ncbi.nlm.nih.gov/9227864/)
11. Frénay HME, Bunschoten AE, Schouls LM, Leeuwen WJ van, Vandenbroucke-Grauls CMJE, Verhoef J, et al. Molecular typing of methicillin-resistant Staphylococcus aureus on the basis of protein A gene polymorphism. *Eur J Clin Microbiol Infect Dis*. 1996; 15: 60–64. PMID: [8641305](https://pubmed.ncbi.nlm.nih.gov/8641305/)
12. Shopsis B, Gomez M, Montgomery SO, Smith DH, Waddington M, Dodge DE, et al. Evaluation of Protein A Gene Polymorphic Region DNA Sequencing for Typing of Staphylococcus aureus Strains. *J Clin Microbiol*. 1999; 37: 3556–3563. PMID: [10523551](https://pubmed.ncbi.nlm.nih.gov/10523551/)
13. Wassenberg MWM, Kluytmans JAJW, Bosboom RW, Buiting AGM, van Elzakker EPM, Melchers WJG, et al. Rapid diagnostic testing of methicillin-resistant Staphylococcus aureus carriage at different anatomical sites: costs and benefits of less extensive screening regimens. *Clin Microbiol Infect*. 2011; 17: 1704–1710. doi: [10.1111/j.1469-0691.2011.03502.x](https://doi.org/10.1111/j.1469-0691.2011.03502.x) PMID: [21595786](https://pubmed.ncbi.nlm.nih.gov/21595786/)
14. Struelens MJ, Hawkey PM, French GL, Witte W, Tacconelli E. Laboratory tools and strategies for methicillin-resistant Staphylococcus aureus screening, surveillance and typing: state of the art and unmet needs. *Clin Microbiol Infect*. 2009; 15: 112–119. doi: [10.1111/j.1469-0691.2008.02188.x](https://doi.org/10.1111/j.1469-0691.2008.02188.x) PMID: [19438632](https://pubmed.ncbi.nlm.nih.gov/19438632/)
15. Manian FA, Senkel D, Zack J, Meyer L. Routine Screening for Methicillin-Resistant Staphylococcus aureus Among Patients Newly Admitted to an Acute Rehabilitation Unit. *Infect Control Hosp Epidemiol*. 2002; 23: 516–519. PMID: [12269449](https://pubmed.ncbi.nlm.nih.gov/12269449/)
16. Reagan DR, Doebbeling BN, Pfaller MA, Sheetz CT, Houston AK, Hollis RJ, et al. Elimination of Coincident Staphylococcus aureus Nasal and Hand Carriage with Intranasal Application of Mupirocin Calcium Ointment. *Ann Intern Med*. 1991; 114: 101–106. PMID: [1898585](https://pubmed.ncbi.nlm.nih.gov/1898585/)



17. Parras F, Guerrero MC, Bouza E, Blázquez MJ, Moreno S, Menarguez MC, et al. Comparative study of mupirocin and oral co-trimoxazole plus topical fusidic acid in eradication of nasal carriage of methicillin-resistant *Staphylococcus aureus*. *Antimicrob Agents Chemother*. 1995; 39: 175–179. PMID: [7695302](#)
18. Koreen L, Ramaswamy SV, Graviss EA, Naidich S, Musser JM, Kreiswirth BN. spa Typing Method for Discriminating among *Staphylococcus aureus* Isolates: Implications for Use of a Single Marker To Detect Genetic Micro- and Macrovariation. *J Clin Microbiol*. 2004; 42: 792–799. PMID: [14766855](#)
19. Harris SR, Feil EJ, Holden MTG, Quail MA, Nickerson EK, Chantratita N, et al. Evolution of MRSA During Hospital Transmission and Intercontinental Spread. *Science*. 2010; 327: 469–474. doi: [10.1126/science.1182395](#) PMID: [20093474](#)
20. Ben-David D, Mermel LA, Parenteau S. Methicillin-resistant *Staphylococcus aureus* transmission: The possible importance of unrecognized health care worker carriage. *Am J Infect Control*. 2008; 36: 93–97. doi: [10.1016/j.ajic.2007.05.013](#) PMID: [18313510](#)
21. Cookson B, Peters B, Webster M, Phillips I, Rahman M, Noble W. Staff carriage of epidemic methicillin-resistant *Staphylococcus aureus*. *J Clin Microbiol*. 1989; 27: 1471–1476. PMID: [2768437](#)
22. Onnela J-P, Christakis NA. Spreading paths in partially observed social networks. *Phys Rev E*. 2012; 85: 036106. PMID: [22587148](#)
23. Temime L, Opatowski L, Pannet Y, Brun-Buisson C, Boëlle PY, Guilletot D. Peripatetic Health-Care Workers as Potential Superspreaders. *Proc Natl Acad Sci*. 2009; 106: 18420–18425. doi: [10.1073/pnas.0900974106](#) PMID: [19841263](#)
24. Hornbeck T, Naylor D, Segre AM, Thomas G, Herman T, Polgreen PM. Using Sensor Networks to Study the Effect of Peripatetic Healthcare Workers on the Spread of Hospital-Associated Infections. *J Infect Dis*. 2012; 206: 1549–1557. doi: [10.1093/infdis/jis542](#) PMID: [23045621](#)
25. Boyce JM, Pittet D. Guideline for Hand Hygiene in Health-Care Settings: recommendations of the Healthcare Infection Control Practices Advisory Committee and the HICPAC/SHEA/APIC/IDSA Hand Hygiene Task Force. *Infect Control Hosp Epidemiol*. 2002; 23: S3–S40. PMID: [12515399](#)
26. Pittet D, Allegranzi B, Sax H, Dharan S, Pessoa-Silva CL, Donaldson L, et al. Evidence-based model for hand transmission during patient care and the role of improved practices. *Lancet Infect Dis*. 2006; 6: 641–652. PMID: [17008173](#)
27. Barclay VC, Smieszek T, He J, Cao G, Rainey JJ, Gao H, et al. Positive Network Assortativity of Influenza Vaccination at a High School: Implications for Outbreak Risk and Herd Immunity. *PLoS ONE*. 2014; 9: e87042. doi: [10.1371/journal.pone.0087042](#) PMID: [24505274](#)
28. Smieszek T, Salathé M. A low-cost method to assess the epidemiological importance of individuals in controlling infectious disease outbreaks. *BMC Med*. 2013; 11: 35. doi: [10.1186/1741-7015-11-35](#) PMID: [23402633](#)
29. Starnini M, Machens A, Cattuto C, Barrat A, Pastor-Satorras R. Immunization strategies for epidemic processes in time-varying contact networks. *J Theor Biol*. 2013; 337: 89–100. doi: [10.1016/j.jtbi.2013.07.004](#) PMID: [23871715](#)
30. Isella L, Romano M, Barrat A, Cattuto C, Colizza V, Van den Broeck W, et al. Close Encounters in a Pediatric Ward: Measuring Face-to-Face Proximity and Mixing Patterns with Wearable Sensors. *PLoS ONE*. 2011; 6: e17144. doi: [10.1371/journal.pone.0017144](#) PMID: [21386902](#)
31. Smieszek T, Barclay VC, Seeni I, Rainey JJ, Gao H, Uzicanin A, et al. How should social mixing be measured: comparing web-based survey and sensor-based methods. *BMC Infect Dis*. 2014; 14: 136. doi: [10.1186/1471-2334-14-136](#) PMID: [24612900](#)
32. Lucet J-C, Laouenan C, Chelius G, Veziris N, Lepelletier D, Friggeri A, et al. Electronic sensors for assessing interactions between healthcare workers and patients under airborne precautions. *PloS One*. 2012; 7: e37893. doi: [10.1371/journal.pone.0037893](#) PMID: [22662245](#)
33. Watts DJ, Strogatz SH. Collective dynamics of “small-world” networks. *Nature*. 1998; 393: 440–442. PMID: [9623998](#)
34. Hollingsworth JC, Chisholm CD, Giles BK, Cordell WH, Nelson DR. How Do Physicians and Nurses Spend Their Time in the Emergency Department? *Ann Emerg Med*. 1998; 31: 87–91. PMID: [9437348](#)
35. Keeling MJ, Eames KTD. Networks and epidemic models. *J R Soc Interface*. 2005; 2: 295–307. PMID: [16849187](#)
36. Bradley SF, Terpenning MS, Ramsey MA, Zarins LT, Jorgensen KA, Sottile WS, et al. Methicillin-resistant *Staphylococcus aureus*: Colonization and Infection in a Long-term Care Facility. *Ann Intern Med*. 1991; 115: 417–422. PMID: [1908198](#)
37. Bilavsky E, Lerman Y, Rabinovich A, Salomon J, Lawrence C, Rossini A, et al. Carriage of methicillin-resistant *Staphylococcus aureus* on admission to European rehabilitation centres—a prospective study. *Clin Microbiol Infect*. 2012; 18: E164–E169. doi: [10.1111/j.1469-0691.2012.03851.x](#) PMID: [22512722](#)

38. Lauderdale T-LY, Wang J-T, Lee W-S, Huang J-H, McDonald LC, Huang I-W, et al. Carriage rates of methicillin-resistant *Staphylococcus aureus* (MRSA) depend on anatomic location, the number of sites cultured, culture methods, and the distribution of clonotypes. *Eur J Clin Microbiol Infect Dis*. 2010; 29: 1553–1559. doi: [10.1007/s10096-010-1042-8](https://doi.org/10.1007/s10096-010-1042-8) PMID: [20820833](https://pubmed.ncbi.nlm.nih.gov/20820833/)
39. Albert R, Barabási A-L. Statistical mechanics of complex networks. *Rev Mod Phys*. 2002; 74: 47–97.



## **S1 Text: I-Bird investigation methods**

### **Design and participation**

The I-Bird investigation was a prospective, observational, cohort study conducted in a 200-bed long-term care hospital over a 6-month period (June–November) in 2009. This center, located in Berck-sur-Mer, France, is organized in five wards corresponding to five different clinical specialties (geriatrics, neurology, nutrition, orthopedics and post-operative cares). Such a facility was chosen for several reasons. These centers are viewed as a potential reservoir of antibiotic-resistant bacteria for acute-care centers, with the potential for amplification and dissemination into the community. Second, the average stay of 3 months provided a longer follow-up of the patients, thereby improving the study feasibility in terms of wireless recording of interactions among participants.

To enhance the participation of patients and healthcare workers (HCWs), several information meetings were held in April and May 2009. The overall participation rate was 90.1% (788 participants total) of all patients and HCW present in the hospital over the 6-month study period: 452 distinct patients stayed in the hospital, 108 auxiliary nurses, 76 nurse interns, 58 nurses, 28 reeducation therapists (including physiotherapists and ergotherapists), 19 ancillary hospital staff, 7 nurse managers, 7 physicians, and 40 administrative staff. Overall, 8,599 participant-weeks of follow-up were recorded. Each participant was followed for an average of 12.5 weeks (7.6 for patients, 15.2 for HCWs) and up to a maximum of 22 weeks.

## Bacterial follow-up and characterization

Microbiological monitoring of a participating patient included weekly nasal and rectal swabs, as well as swabs of entry wounds or invasive devices (tracheotomy, gastrostomy, *etc.*) for *Staphylococcus aureus* and enterobacteria resistant to third-generation cephalosporins. Screening of participating HCWs included only weekly nasal swabs. All collected swabs (nasal, rectal, and entry wounds combined) were frozen at -80°C and stored in the Microbiological Laboratory of Raymond Poincaré University Hospital (Garches, France). The targeted microorganisms were isolated, subjected to antibiotic-resistance phenotyping and genotyped (to date only *S. aureus*).

## Microbiological methods for *S. aureus* identification and characterization

Nasal swabs collection was performed as described previously with sterile cotton-wool swabs[1]. They were immediately placed in transport medium and sent to the Microbiology Laboratory where they were inoculated into 500 µl of brain-heart infusion medium. Then, 100 µl of this broth were plated onto Chapman agar (Biomérieux®, Marcy l'Etoile, France) and MRSA ID screening agar (Biomérieux®) and incubated for 48 h at 36°C. *S. aureus* was identified by mass spectrometry using the MALDI-TOF-MS spectrometer with flex control software (Autoflex; Bruker Daltonics, Bremen, Germany)

Antimicrobial susceptibility of all isolated *S. aureus* strains was determined on Mueller-Hinton agar plates (BioMerieux®) using the disk diffusion method[2]. The screening procedure had an expected sensitivity of 61.5% and specificity of 98.8%, [3] although higher sensitivity value has been reported (~80%). [4] Twenty antibiotics were tested, including methicillin, which was subsequently used to classify strains as methicillin-resistant (MRSA) or

methicillin-sensitive (MSSA). Antibiotic resistance was described as Sensitive-Intermediate-Resistant.

To describe *S. aureus* clonal diversity, all strains isolated from the first positive nasal swab were *spa* typed by Genoscreen (Lille, France) as previously described[5,6]. Briefly, Polymerase Chain Reaction used primers 1113F (TAAAGACGATCCTTCGGTGTGAGC) and 1496R (TTTGCTTTTGCAATGTCATTTACTG) in the following cycle conditions: 12 min at 95°C, 40 cycles: 30 s at 95°C, 30 s at 62°C, 30 s at 72°C, 10 min at 72°C and 15 min at 15°C. DNA sequences were obtained with a 3730xl DNA Analyzer (Applied Biosystems). *Spa* types and minimum spanning tree were determined with Bionumerics v6.5 software (Applied Maths, Saint Martens, Belgium). A broad diversity of *spa* types were isolated, denoting a high variability in *S. aureus* strains (Fig. 1).

### Close proximity interaction recording

We obtained large-scale dynamic networks of interactions among patients and HCWs within the hospital. These interactions were measured by equipping all study participants with a “LogSensor” (WSN430): a wireless device to signal and record the presence of others LogSensors devices in their short radio-frequency range. Our team developed the LogSensor hardware and software. LogSensor is powered by a TI MSP430 micro-controller and communicates through a TI Chipcon CC1100 packet radio interface. The sensor design is classical (see Fig. 2). Interactions between two sensors are recorded by broadcasting periodical beacons (“Hello” packets). When a node *X* received a “Hello” packet from another node *Y*, it logged that at a *Y* to *X* interaction had occurred at local time *t*. The software was implemented as a loop performing simple operations: (i) periodically sending/receiving “Hello” packets; (ii) uploading the data when the memory was almost full. The most

restrictive element was battery life. Therefore, sensors were kept in a low-power mode (sleep mode) for most of the time, and only awakened to send and receive information (Fig. 3). Using a sleep period extended the battery longevity but also increased the probability of missing incoming “Hello” packets. Therefore the software allowed for discovery of all nodes that remained at a distance  $< 1.5$  m during a given period, and for those trying to send their beacons at the same time.

### Deployment details

LogSensors were distributed in batches during the first 5 weeks of the investigation (1 week for each ward). All participants within the hospital (medical & administrative staff, patients) were asked to keep their LogSensors with them continuously during their presence in the hospital. For HCWs, the sensor was kept in the overcoat pocket. For patients, it could be in a pocket, worn as a watch or around an ankle and was placed at bedside during the night. For patients unable to move, it was attached to the bed.

Five hundred and eleven LogSensors were distributed during the investigation period and 817 participant files were recorded. Each participant’s file contained data for up to 22 weeks of follow-up. During the study period, recording was discontinued on three occasions for 1 or 2 days to change the sensor batteries. In case of battery failure between changes, the interactions were not recorded for this participant until next scheduled battery change.

### LogSensor design, calibration and management

At the start of the experiment, each LogSensor’s memory was empty and a sensor only knew its unique identifier. Thereafter, sensors distributed to all experiment participants periodically (every 30 s) recorded their surrounding neighborhood. We expected recordings

of close interactions, i.e., LogSensors within 1.5-m radius. From experiments made using such sensors, in a clean environment (line of sight), the recordings can vary significantly between two sensors due to hardware-components variations and tolerance. For example, in some experiments using a set of 3 equidistant sensors (60 cm), received signal-strength indication (RSSI) ranged from  $\sim$ -57 dBm to  $\sim$ -47 dBm. In the i-Bird experiment, sensors were carried by participants and body-water content was able to efficiently block the 2.4GHz signal, thus allowing the recording of face-to-face interactions only. Because we are interested in close contact, we concentrated only on high-RSSI signals (Fig. 4).

We chose to record the mean RSSI of all consecutive “Hello” packets received from a given sensor. All the data recorded were saved in the sensors’ onboard memory, meaning that no one could get all the data without having physical access to the sensors during the experiment. To sustain a 6-month-long experiment, data were compressed: instead of recording all contacts one by one, consecutive recordings between two sensors were grouped, allowing one packet loss and keeping only the mean RSSI of all consecutive packets.

To ensure experiment parameters, the LogSensors could also be controlled wirelessly by a management PC. This additional feature cannot be used to retrieve the recorded data but to monitor the good functioning of the sensors that crossed paths during the day (battery levels, anomalies). This information was not recorded and only the last “Hello” packet transmission was used to detect faulty sensors. Management PCs were deployed in specific hospital rooms and only required access to the local area network to exchange data with each other.

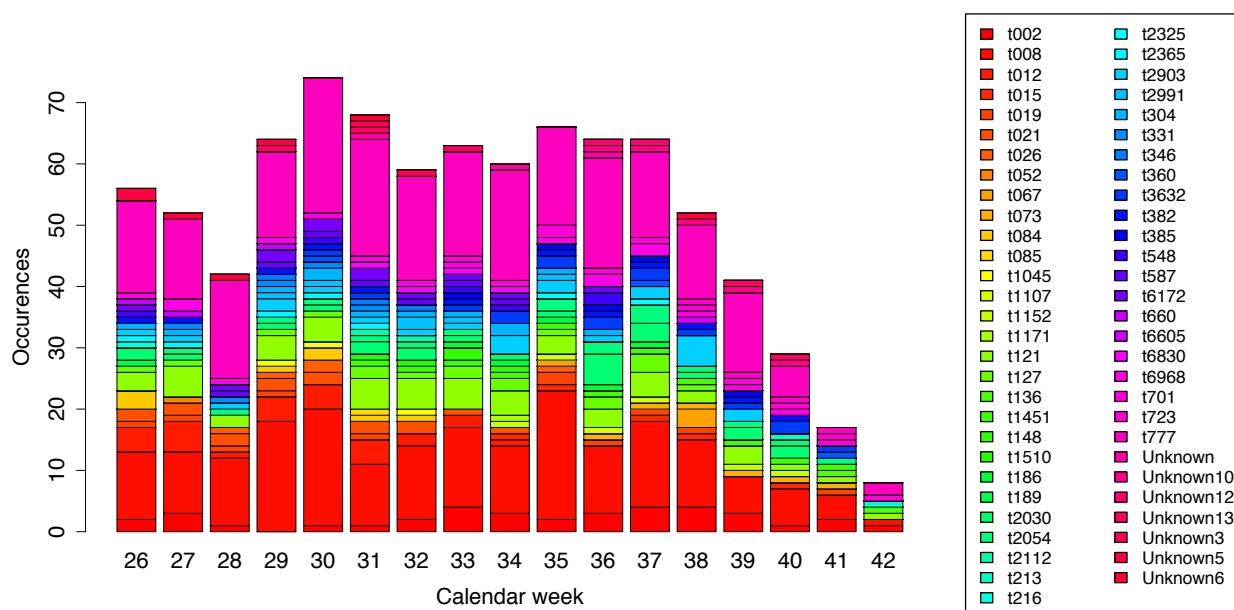
## References

1. Alvarez AS, Remy L, Allix-Béguec C, Ligier C, Dupont C, Leminor O, et al. Patient nostril microbial flora: individual-dependency and diversity precluding prediction of *Staphylococcus aureus* acquisition. *Clin Microbiol Infect*. 2014;20: 70–78. doi:10.1111/1469-0691.12208
2. Leclercq R, Cantón R, Brown DFJ, Giske CG, Heisig P, MacGowan AP, et al. EUCAST expert rules in antimicrobial susceptibility testing. *Clin Microbiol Infect*. 2013;19: 141–160. doi:10.1111/j.1469-0691.2011.03703.x
3. Wassenberg MWM, Kluytmans JAJW, Bosboom RW, Buiting AGM, van Elzaker EPM, Melchers WJG, et al. Rapid diagnostic testing of methicillin-resistant *Staphylococcus aureus* carriage at different anatomical sites: costs and benefits of less extensive screening regimens. *Clin Microbiol Infect*. 2011;17: 1704–1710. doi:10.1111/j.1469-0691.2011.03502.x
4. Struelens MJ, Hawkey PM, French GL, Witte W, Tacconelli E. Laboratory tools and strategies for methicillin-resistant *Staphylococcus aureus* screening, surveillance and typing: state of the art and unmet needs. *Clin Microbiol Infect*. 2009;15: 112–119. doi:10.1111/j.1469-0691.2009.02698.x
5. Frénay HME, Bunschoten AE, Schouls LM, Leeuwen WJ van, Vandenbroucke-Grauls CMJE, Verhoef J, et al. Molecular typing of methicillin-resistant *Staphylococcus aureus* on the basis of protein A gene polymorphism. *Eur J Clin Microbiol Infect Dis*. 1996;15: 60–64. doi:10.1007/BF01586186
6. Shopsin B, Gomez M, Montgomery SO, Smith DH, Waddington M, Dodge DE, et al. Evaluation of Protein A Gene Polymorphic Region DNA Sequencing for Typing of *Staphylococcus aureus* Strains. *J Clin Microbiol*. 1999;37: 3556–3563.



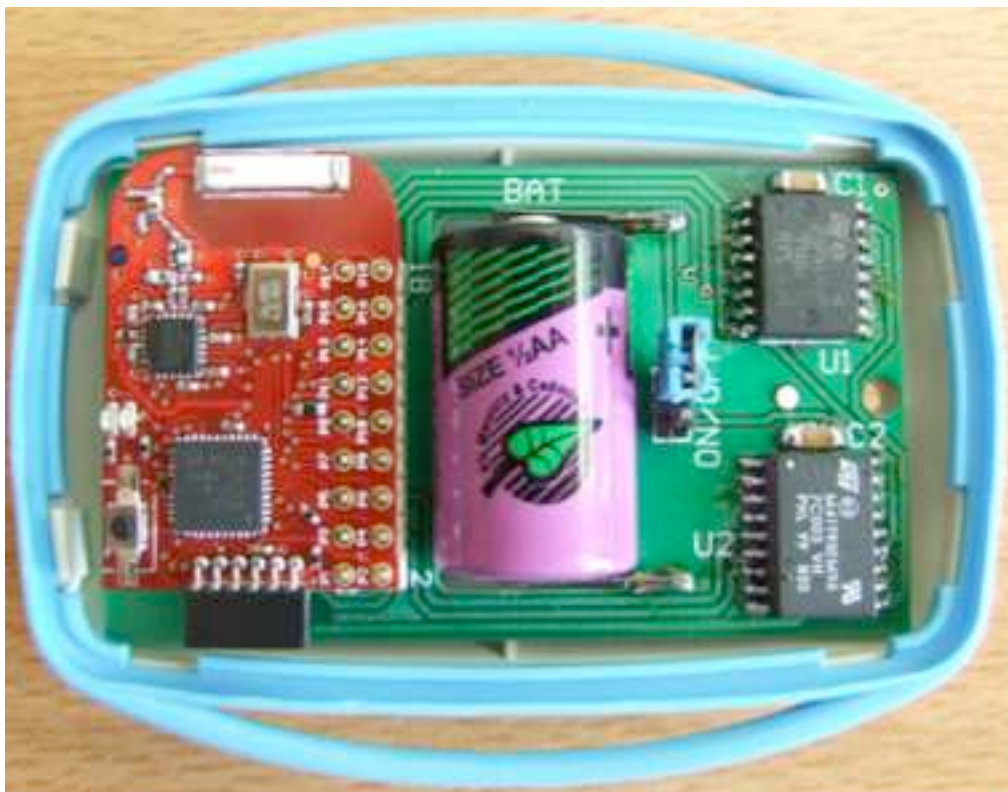
## Figures

**Figure 1 – Numbers of *spa* types isolated from patients as a function of time in the hospital.** The decrease in the last 4 weeks corresponds to the progressive stop in systematic swabbing due to patient discharge without replacement.

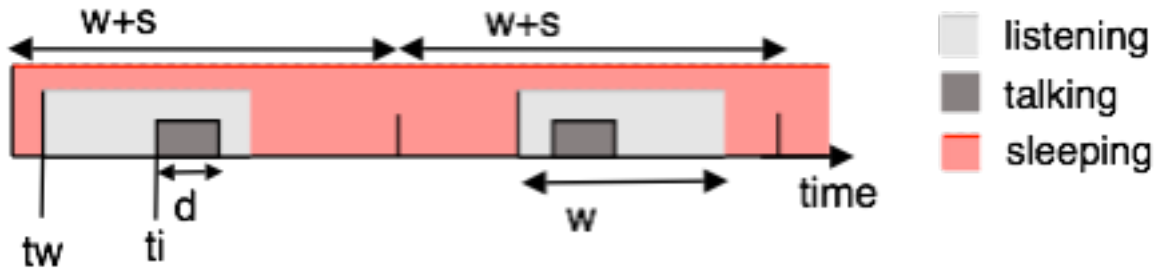




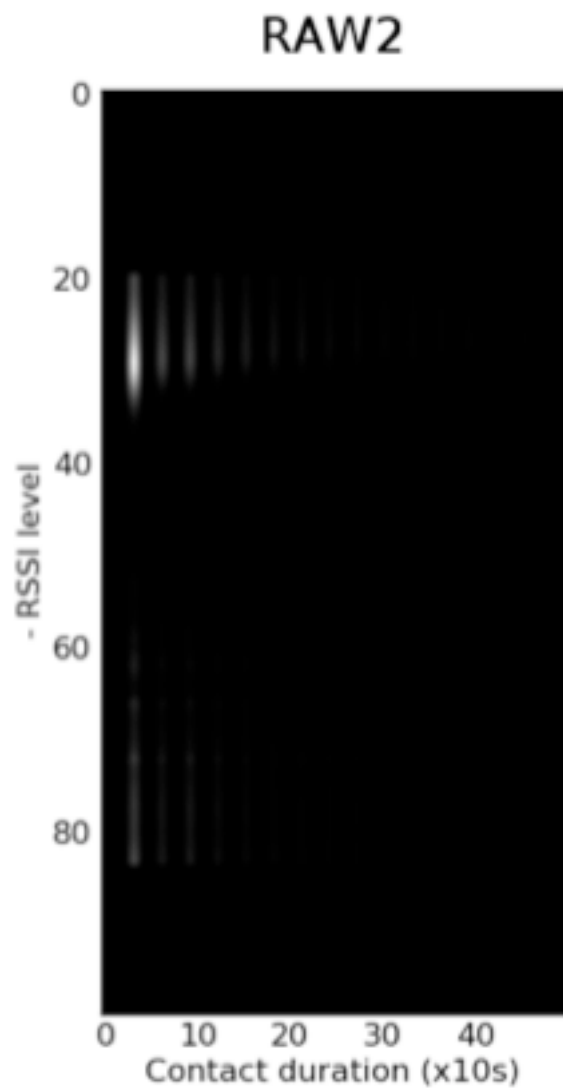
**Figure 2 – The LogSensor design for the i-Bird experiment.** We used the eZ430-RF2500 board provided by TI (red), that incorporates a highly integrated ultra-low-power MSP430 MCU with 16-MHz performance (MSP430F2274), a 2.4-GHz wireless transceiver (CC2500) and a chip antenna. We added on our own board a 64-Megabytes serial flash memory and a real time clock with a built-in 32,768 kHz oscillator. The sensor is powered by a 3.6 V primary lithium-thionyl chloride (Li-SOCl<sub>2</sub>) high-energy density 1/2 AA-size battery.



**Figure 3 – « Hello » protocol with sleep period.** Each sensor can be in one of the following three states: listening, talking or sleeping. These three states are present inside a timeframe  $F$ , of duration  $w + s$ . During  $F$ , a sensor randomly picks an instant  $t_w$ , such that  $t_w \in [0 ; s]$ . Then, the sensor randomly selects a second instant  $t_i$ , such that  $t_i \in [t_w ; t_w + w - d]$ . The sensor is in talking state during  $[t_i ; t_i + d]$ , in listening state during  $[t_w ; t_w + w] \setminus [t_i ; t_i + d]$  and in sleeping state the rest of the timeframe  $F$ . The “Hello” message is transmitted at  $t_i$  with duration  $d$ . For each timeframe  $F$ , the node transmits only one message of duration  $d$ , listens to its surroundings during  $w - d$  and sleeps during  $s$ . The medium access probability is therefore:  $P = d / (w + s)$ .



**Figure 4 – Filtering based on the distribution of mean RSSI signals as a function of contact duration.** Since the wireless channel is very versatile, we will only consider packets having a strong RSSI ( $> -40$  dBm). All other packets (gray line near  $-80$  dBm) will be considered as noise.





## S2 Text: Description of the dynamic CPI network

### Supplementary Methods: building the dynamic CPI network

Individual sensor records were extracted and provided raw contact data. The raw data was cleaned by dropping low-intensity signals, removing parasite records using HCWs work schedules and symmetrizing records to correct for occasional recording glitches. Finally, we transformed the consecutive records with the same sensor in start/duration data, and cumulated the time of interaction between pairs by day. This led to the definition of a dynamic network of 120 successive daily CPI networks, where nodes were the participants and an edge existed when a CPI lasting  $\geq 30$  s existed during a day. For each patient admitted during the study period, we recorded the number of days to the first direct CPI with others in the hospital. The cumulative number of those neighbors in the CPI network was computed as a function of time since admission, as well as that of 2-hop neighborhoods (i.e. neighbors of neighbors) and 3-hop neighborhoods (see Fig. 5).

The percentage of between-wards CPIs was compared to that expected if the same CPIs had been random between all participants using a Chi-square test.

The daily numbers of patients' CPIs during the first month after admission was studied by using a random-effects linear model, including a random effect for date and individual according to the equation  $Y_{it} = \mu + Id_i + D_t + e_{it}$ , where  $Y_{it}$  was the number of CPIs for individual  $i$  on day  $t$ ,  $\mu$  the average number of CPIs,  $Id_i$  the individual-specific random effect,  $D_t$  the day and  $e_{it}$  an error term. HCWs' CPIs were investigated using a similar random-effects linear model where  $D_t$

corresponded to the day of the week. The existence of a between-individual variability was tested by comparing the between-subjects variance to 0 with a restricted likelihood ratio test.[1]

### Dynamics of the CPI network

The dynamic CPI network among individuals in the hospital was dense: a CPI occurred in 19.3% of the 120,735 potential daily interactions between pairs of individuals. Most contacts occurred within wards (81.7% CPIs within wards vs. 29.5% expected by chance,  $P < 0.0001$ ; Fig. 1 in main text), yet most participants also had CPIs in another ward during their time in-hospital (97%, 481 out of 495 affected to a ward). CPIs outside the ward were highly heterogeneous: 80% of all between-wards CPIs were attributable to only 40% of all participants. The rate of outside-the-ward CPIs was higher for patients than HCWs (2.8 vs. 2.1/person per day,  $P = 0.016$ ).

The degree distribution in the daily CPI networks was right-skewed, indicating that while most individuals had CPIs with a limited number of neighbors on each day, a few had a large number of CPIs (Fig. 6). On a given day, the shortest path between any two individuals had  $5 \pm 3$  (mean  $\pm$  SD) intermediaries, much fewer than the number of individuals in the network. Patients had CPIs with  $12 (\pm 6.2)$  distinct persons each day, half of them with other patients ( $n = 6 \pm 4.5$ ) and the others with HCWs ( $n = 6 \pm 2.5$ ); whereas HCWs had CPIs with  $15 (\pm 7.2)$  persons, among whom  $9 (\pm 5.7)$  were patients and  $6 (\pm 3.1)$  other HCWs. The corresponding median numbers of CPIs were 9 (range: [1-98]) for patients and 12 (range [1-84]) for HCWs. While the numbers of CPIs were within the same range for patients and HCWs, the daily-cumulative durations of the CPIs were much longer for patients than HCWs, respectively:  $12.2 (\pm 11.3)$  h vs.  $3.7 (\pm 2.4)$  h (Fig.

2 in main text). Defining a participant's  $n$ -hop neighborhood as all individuals within  $n$  steps from him revealed that a majority of people in the hospital was within the 3-hop neighborhood of any participant, another illustration of the small-world characteristics of the CPI network (Fig. 5).

Patients with the highest CPI frequencies tended to have more CPIs than other patients every day: the between-patient variability accounted for 36% of the total CPI number variance, more than the day-to-day variation for the same patient (25% of the variance). Likewise, the cumulative CPI durations varied widely among participants, with coefficients of variation (standard deviation/mean) of 92% for patients and 66% for HCWs. The number of distinct persons with whom CPIs occurred increased quickly with the time since admission, reaching 50% of its final expanse after only 5 days (Fig. 5). However, these values varied markedly among patients, both regarding the final number of neighbors (3 to 246) and the time to acquire them (0 to 108 days). Each patient's extended network, including 2-hop (neighbors of neighbors) and 3-hop neighborhoods, grew accordingly.

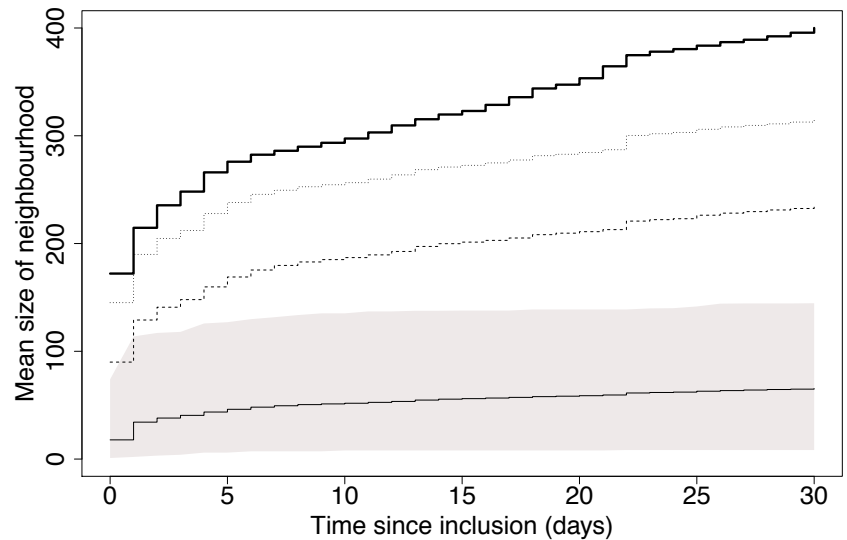
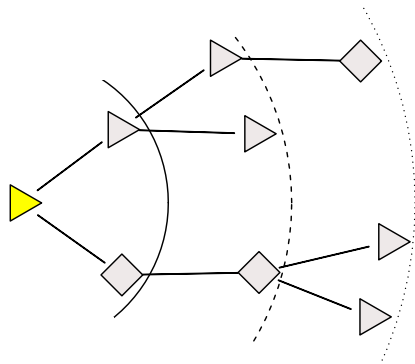
## References

1. Scheipl F, Greven S, Küchenhoff H. Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. *Comput Stat Data Anal.* 2008;52: 3283–3299. doi:10.1016/j.csda.2007.10.022

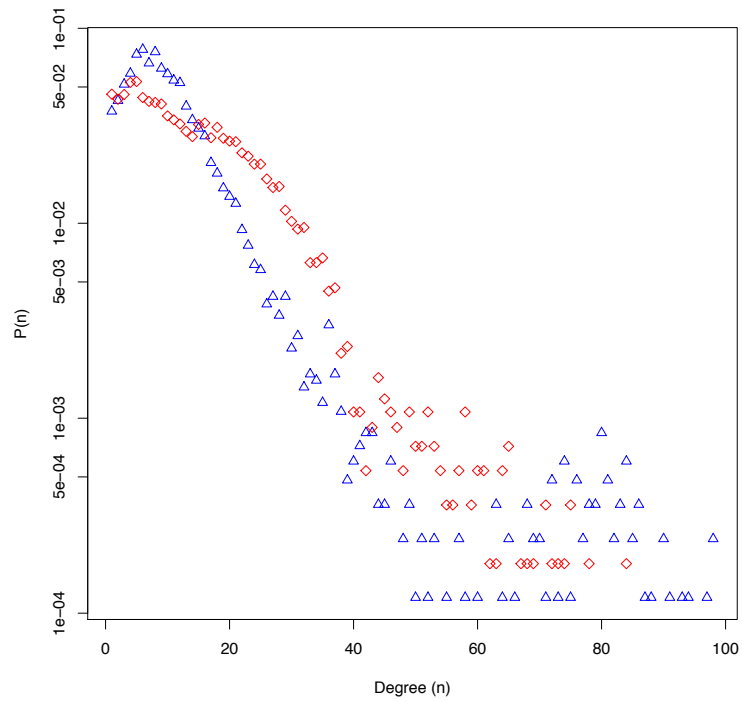


## Figures

**Figure 5 – Change of patients’ neighborhood during the first month after inclusion. (Left)**  $n$ -hop neighborhood includes all individuals whose shortest CPI path from the *ego* (yellow) is smaller or equal to  $n$ . Lines show the 1-hop neighborhood (solid), 2-hop (dashed) and 3-hop (dotted). Patients are shown as triangles and HCWs as diamonds. **(Right)** Mean size of patients’ neighborhoods during the first month of hospitalization (1-hop: solid, 2-hop: dashed, 3-hop: dotted, maximum: bold). The maximum neighborhood includes all individuals that could have been contacted on each day. The gray area limits the 2.5% to 97.5% percentile of the 1-hop neighborhood size distribution.



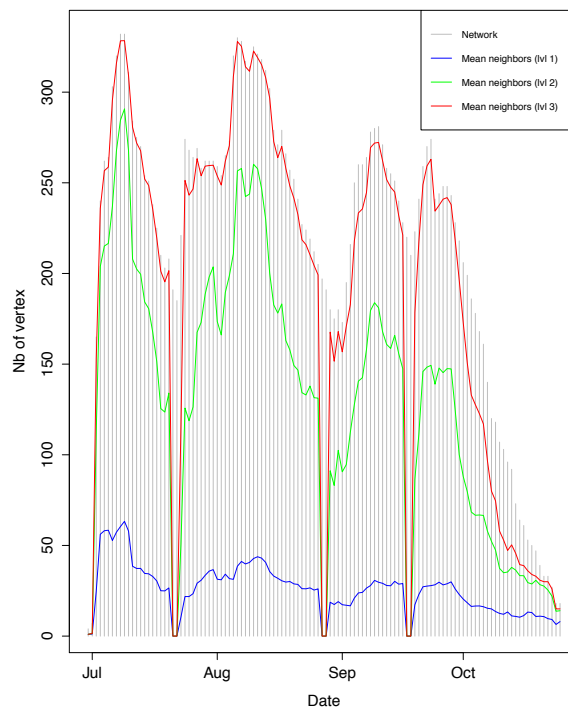
**Figure 6 – Degree distribution in patients and HCWs.** Patients are shown as blue triangles and HCWs as red diamonds.



### 3.4 Commentaires et perspectives

#### 3.4.1 Sur les conséquences d'un réseau « petit-monde » dynamique

Dans le chapitre 2, nous avons remarqué que le réseau des contacts dans I-Bird présentait des propriétés petit-monde. Cette particularité rend l'existence d'un chemin qui relie deux nœuds tirés au hasard très probable. Le passage des graphes quotidiens au réseau agrégé temporellement cohérent ne fait qu'augmenter ce phénomène. En conséquence, la stratégie (S1) définie dans l'article ne permet pas de tester la corrélation entre contacts et portage de *S. aureus* : même dans le cadre de l'hypothèse nulle, la percolation dans le réseau permettra tout de même de trouver un chemin reliant un cas incident à un autre individu ayant porté la même souche à l'échantillonnage précédent. Ce phénomène est illustré par la Figure 23 : à l'image de la Figure 19, nous pouvons représenter le voisinage  $n$ -hop moyen d'un participant, en utilisant cette fois des graphes calculés de façon à retranscrire le voisinage temporellement cohérent au cours des 7 jours précédents.



**Figure 23. Nombre des voisins des participants de l'étude I-Bird avec un effet mémoire.** Taille moyenne du voisinage d'un participant en fonction de la date, avec agrégation du réseau égo-centré remontant jusqu'à une semaine. Le voisinage direct est représenté en bleu, le voisinage 2-hop en vert et 3-hop en rouge. Les barres grises représentent le nombre total de capteurs actifs chaque jour.

### **3.4.2 Sur l'existence du portage transitoire**

La granularité des données de portage fait qu'un individu peut être colonisé de façon transitoire, et décolonisé avant le prélèvement suivant : le suivi longitudinal ne permet pas de capter ces épisodes. Ce comportement a été documenté chez les personnels de santé,[3, 14, 28, 35] et s'explique en partie par une application imparfaite des mesures d'hygiène. Le niveau d'adhérence aux mesures d'hygiène est lui-même corrélé au type de contacts entre patients et soignants (notamment aux soins prodigués).[50, 95] C'est pour cette raison que nous n'avons investigué que les épisodes de colonisation chez des patients. Ces porteurs transitoires ne remplissent pas le critère de portage pour pouvoir être identifiés comme « transmitter candidates » dans l'article : ils ne pourront pas être considérés comme source probable de l'infection pour un cas incident. La stratégie (S2) proposée dans cet article ignore ce type de portage, en ne s'intéressant qu'au pourcentage de cas incidents où un infecteur potentiel est trouvé en contact direct avec le cas. La stratégie (S3) s'intéresse enfin à la longueur (en nombre d'intermédiaires) du chemin le plus court permettant de relier un cas incident à ses infecteurs potentiels : par construction, si l'infecteur n'est pas en contact direct avec le cas incident, les intermédiaires ne peuvent pas être porteurs de la même souche (sinon, ils auraient été préférés comme « CPI-supported transmitters », car plus proches du cas incident sur le réseau). Ces individus, bien qu'apparaissant non colonisés, ont très probablement joué un rôle dans la transmission, à l'interface entre une source de *S. aureus* et un autre patient susceptible.

Dans l'hypothèse (irréaliste) où la granularité des données de portage serait aussi fine que celle des données de contacts, les chemins de colonisation seraient observables en temps réel, et les épisodes d'incidence se produiraient nécessairement lors d'un contact avec une personne colonisée. Avec la stratégie adoptée dans l'article, nous proposons une méthodologie qui tient compte de cette différence de résolution en autorisant des chaînes de transmission où les intermédiaires ne sont pas colonisés lors de l'échantillonnage.

### **3.4.3 Sur la réalité des chaînes de transmission**

Les données de colonisation, en plus d'être espacées d'une semaine chez un individu donné, sont sujettes à la sensibilité et la spécificité de la procédure de détection. En particulier, chaque test négatif a une probabilité non négligeable d'être un faux négatif

(entre 20% et 38.5%). Ceci nous amène à manquer certains phénomènes de portage, qui viennent s'ajouter au portage transitoire non observé. De plus, la granularité des données fait qu'une chaîne de transmission peut être interprétée de façon erronée : supposons un individu A porteur à la date  $d$ , et deux individus B et C porteurs à la date  $d+\tau$ . Si la chaîne réelle est  $A \rightarrow B$  (date  $d_1 \in [d; d + \tau]$ ) puis  $B \rightarrow C$  (date  $d_2 \in ]d_1; d + \tau[$ )), il est possible selon les dates de prélèvement et le réseau observé que les deux épisodes d'incidence chez B et C soient imputés à A. La réalité des chaînes de transmission est à nuancer, mais la procédure que nous développons permet d'identifier des liens entre réservoir d'infection et individu néo-colonisé.

#### **3.4.4 Sur la permutation du portage pour simuler l'hypothèse d'indépendance**

Deux approches auraient pu être envisagées pour simuler la distribution des différentes statistiques proposées sous l'hypothèse d'indépendance entre contacts portage.

Celle que nous avons adoptée permet de conserver la structure observée du réseau. En particulier, nous conservons le clustering des 5 services et la topologie générale du réseau, qui retranscrit l'organisation des équipes et le fonctionnement de l'hôpital étudié. Lors des permutations, l'autocorrélation longitudinale est conservée, et comme la prévalence de *S. aureus* était similaire chez les patients et le personnel de santé, nous n'avons pas tenu compte de ce statut pour les répartitions. Nous conservons également la distribution du nombre de chaque souche isolée, qui reflète la pression de colonisation chaque semaine, pour chaque souche.

Une autre approche envisageable aurait été de conserver le portage individuel, mais de réorganiser le réseau. Toutefois, cette approche est beaucoup plus difficile à mettre en œuvre, car elle nécessite de modéliser les processus qui amènent à la formation (et dissolution) de liens, *i.e.* les processus sociaux qui sont à l'origine du réseau de contacts.

#### **3.4.5 Sur le modèle de dissémination dans les simulations**

Le modèle de dissémination utilisé est volontairement simplifié : notamment, il ne tient pas compte de la durée des contacts pour pondérer la probabilité de transmission. Ces probabilités ont été calculées afin de refléter la prévalence moyenne observée dans l'étude I-Bird. De plus, il est intéressant de noter que dans ces simulations, le pathogène ne peut être transmis qu'aux contacts directs au cours d'un pas de temps (1 jour). Le portage consécutif entre plusieurs individus ayant des contacts successifs n'est donc pas simulé : ceci explique en partie l'écart relatif entre la proportion d'infecteurs situés en

contact direct avec leur infecté dans les simulations et permutations (Figure 3 dans l'article).

### **3.4.6 Application pratique**

Le résultat majeur de cet article est la confirmation que les contacts, tels que mesurés par des capteurs de proximité, sont un bon proxy pour évaluer les chemins de dissémination d'une bactérie se transmettant principalement lors des contacts physiques entre deux personnes. Nous validons ainsi l'étude des réseaux de contacts à haute résolution pour l'évaluation de tels chemins : ces réseaux peuvent servir à identifier des situations de risque accru pour l'acquisition de bactéries nosocomiales, et donc être utilisés pour développer des stratégies de contrôle ciblées.

\* \* \*

Nous avons présenté dans ce chapitre une méthode permettant de tester la corrélation entre des données de contact hautement dynamiques et des données de suivi longitudinal de colonisation bactérienne. Nous souhaitons utiliser ces données pour informer les modèles de transmission sur la structure des contacts, et sur les risques individuels que ces contacts peuvent générer. Pour cela, nous allons maintenant détailler une analyse quantitative de ce jeu de données.

## **4 Chapitre 4 : Evaluation quantitative des risques liés au réseau de contacts pour la colonisation des patients par *S. aureus* résistant à la méticilline**

Dans les chapitres précédents, nous avons détaillé l'intérêt de décrire l'hétérogénéité des contacts dans la population, à des fins de modélisation épidémiologique. Une des difficultés dans cette démarche est de définir un « contact », condition nécessaire à l'établissement d'un graphe. Dans cette optique, nous avons démontré que les interactions enregistrées par des capteurs électroniques de proximité fournissaient un proxy adéquat pour les contacts permettant la dissémination de *S. aureus*. Nous allons maintenant utiliser ces données pour quantifier les risques de colonisation par *S. aureus* en milieu hospitalier. Nous nous intéresserons particulièrement aux risques liés au réseau des contacts interindividuels, qui pourrait être à l'origine de nouvelles mesures de prévention et de contrôle dans la lutte contre la dissémination des infections nosocomiales.

### **4.1 Facteurs de risques investigués et définition de la population**

Nous avons réalisé une étude cas-témoins afin de tester l'association de plusieurs facteurs de risques avec la colonisation par *S. aureus*.

#### **4.1.1 Types de facteurs de risques**

L'étude I-Bird étant prospective et observationnelle, les seules données disponibles (en plus des contacts et du suivi longitudinal du portage de *S. aureus*) étaient celles collectées dans le cadre normal du suivi des patients. Nous avons donc envisagé trois catégories de facteurs de risques :

- Les facteurs individuels, reflétant les caractéristiques démographiques des patients ainsi que le niveau de soins qu'ils requièrent, leur historique de colonisation par *S. aureus*, et leur exposition aux antibiotiques,
- Les facteurs liés au seul réseau des contacts des patients avec d'autres individus. Nous avons notamment testé l'influence de plusieurs quantités mesurables sur les réseaux de contacts (comme détaillées à la section 1.4.1.3), certaines ayant déjà été identifiées comme modulant le risque de colonisation,[23, 140]

- Les facteurs associant le réseau des contacts des patients avec les informations de colonisation de leurs contacts.

#### **4.1.2 Choix des cas et des témoins**

Comme dans l'article présenté au Chapitre 3, nous n'analysons que les épisodes d'incidence survenant chez des patients, car le régime de colonisation des soignants est beaucoup plus variable.

Un cas est donc un patient incident pour une souche de *S. aureus* une semaine donnée, c'est à dire que son suivi montre qu'il a été colonisé par la bactérie entre son prélèvement précédent (négatif) et le prélèvement actuel (positif). En cas de portage multiple, nous vérifions que la souche incidente n'était pas présente au cours des deux précédents écouvillonnages, afin de tenir compte de possibles phénomènes de compétition.[33, 81]

Les témoins sont choisis parmi les patients non-colonisés, et appariés sur la date d'incidence afin de refléter la pression de colonisation de chaque souche dans l'hôpital.[70]

### **4.2 Analyse cas-témoins des facteurs de risque de colonisation par SARM**

#### **4.2.1 Résumé de l'article en français**

Le réseau des contacts interindividuels de l'étude I-Bird capture correctement les contacts pouvant mener à la transmission de *S. aureus*. De précédentes études théoriques ont par ailleurs testé avec succès des stratégies de contrôle, lors de la simulation d'une épidémie se propageant sur des réseaux de contacts similaires.[130, 133] D'autres études, plus descriptives, mettent en évidence que certains individus sont plus à risques, en fonction de contacts qu'ils engagent.[23, 140] Ces constats permettent de proposer l'hypothèse que le réseau des contacts permet d'identifier des situations de risque accru chez certains individus, et donc de développer des stratégies d'intervention ciblées, informées par ce type de données. La facilité croissante de collecter de tels enregistrements permet d'envisager un nouvel axe pour lutter contre la dissémination des bactéries nosocomiales (et donc limiter la prolifération de souches multirésistantes).

Nous réalisons une étude cas-témoins où, classiquement, les propriétés des patients incidents sont comparées à celles des patients non-incidents. Nous distinguons trois



catégories de facteurs de risques (individuels, liés au réseau seulement, et liés au réseau enrichi des données de colonisation). Nous comparons les sous-graphes égocentriques, agrégés sur une semaine, centrés sur les cas ou témoins, et se limitant à leur voisinage 1-hop (*i.e.* voisins directs). Les cas et les témoins sont appariés sur la date de colonisation de chaque cas, avec un ratio de 1:8 lorsque c'est possible. Au total, 144 épisodes de colonisation à SARM sont identifiés chez les patients, dont 101 sont inclus dans l'étude. Parallèlement, nous avons testé l'influence de la durée des contacts individuels (avant agrégation quotidienne, puis hebdomadaire) sur les mêmes facteurs de risque. Plus précisément, nous avons retiré successivement du réseau les contacts de moins de  $d$  secondes, et répété l'analyse principale.

Nous identifions dans l'analyse des résultats des facteurs de risque déjà documentés dans la littérature,[31, 136] notamment l'exposition antibiotique des patients, qui favorise la colonisation par SARM (OR = 3.01 (95%CI [1.63–5.59]),  $P=0.0005$ ). Les patients les plus dépendants sont également plus à risque (OR  $\sim 3$ ). Le rôle déjà évoqué du personnel de santé est mis en évidence, avec des contacts plus à risques lorsque les contacts sont avec des personnels de santé. Par contre, les mesures tenant compte du réseau dans son ensemble (comme la centralité ou la densité) ne permettent pas d'identifier des situations ayant un risque significativement différent entre cas et témoins. Enfin, l'analyse conjointe des contacts et du portage des voisins révèle des associations renforcées, dans le même sens que celles impliquant uniquement le réseau.

#### **4.2.2 Article : Interindividual Contacts and Carriage of Methicillin-Resistant *Staphylococcus aureus* : A Nested Case-Control Study, Obadia T. et al., Infection Control and Hospital Epidemiology, 2015**



## ORIGINAL ARTICLE

# Interindividual Contacts and Carriage of Methicillin-Resistant *Staphylococcus aureus*: A Nested Case-Control Study

Thomas Obadia, MSc;<sup>1,2</sup> Lulla Opatowski, PhD;<sup>3,4,5</sup> Laura Temime, PhD;<sup>6</sup> Jean-Louis Herrmann, PharmD, PhD;<sup>7,8</sup> Éric Fleury, PhD;<sup>9,10,\*</sup> Pierre-Yves Boëlle, PhD;<sup>1,2,11,\*</sup> Didier Guillemot, MD, PhD<sup>3,4,5,12,\*</sup> on behalf of the i-Bird Study Group<sup>†</sup>

**BACKGROUND.** Reducing the spread of multidrug-resistant bacteria in hospitals remains a challenge. Current methods are screening of patients, isolation, and adherence to hygiene measures among healthcare workers (HCWs). More specific measures could rely on a better characterization of the contacts at risk of dissemination.

**OBJECTIVE.** To quantify how close-proximity interactions (CPIs) affected *Staphylococcus aureus* dissemination.

**DESIGN.** Nested case-control study.

**SETTING.** French long-term care facility in 2009.

**PARTICIPANTS.** Patients (n = 329) and HCWs (n = 261).

**METHODS.** We recorded CPIs using electronic devices together with *S. aureus* nasal carriage during 4 months in all participants. Cases consisted of patients showing incident *S. aureus* colonization and were paired to 8 control patients who did not exhibit incident colonization at the same date. Conditional logistic regression was used to quantify associations between incidence and exposure to demographic, network, and carriage covariables.

**RESULTS.** The local structure of contacts informed on methicillin-resistant *S. aureus* (MRSA) carriage acquisition: CPIs with more HCWs were associated with incident MRSA colonization in patients (odds ratio [OR], 1.10 [95% CI, 1.04–1.17] for 1 more HCW), as well as longer CPI durations (1.03 [1.01–1.06] for a 1-hour increase). Joint analysis of carriage and contacts showed increased carriage acquisition in case of CPI with another colonized individual (OR, 1.55 [1.14–2.11] for 1 more HCW). Global network measurements did not capture associations between contacts and carriage.

**CONCLUSIONS.** Electronically recorded CPIs inform on the risk of MRSA carriage, warranting more study of in-hospital contact networks to design targeted intervention strategies.

*Infect. Control Hosp. Epidemiol.* 2015;00(0):1–8

Nosocomial transmission of germs is a major public health issue.<sup>1</sup> Because most germs may be transmitted by direct contact, reducing the per-contact risk by improving hygiene has been at the heart of prevention strategies.<sup>2,3</sup> To further reduce risk, it may be necessary to alter the structure of contacts itself, as proposed in theoretical studies.<sup>4–6</sup>

Yet, contacts leading to nosocomial transmission are poorly characterized in hospitals: they are often assumed to occur at

random within the wards,<sup>7</sup> or, at best, equated with attendance of the same ward at the same date among carriers of the same strains.<sup>8,9</sup> Indeed, no convenient way to record contacts existed before advances in communication technologies made it possible to record high-resolution close-proximity interaction (CPI) networks.<sup>10</sup> CPIs correlate with contacts measured in more traditional ways<sup>11</sup> and are not affected by the reporting bias inherent to self-reporting.<sup>12</sup> They may therefore

Affiliations: 1. Sorbonne Universités, UPMC Univ Paris 06, UMR\_S 1136, Institut Pierre Louis d'Epidémiologie et de Santé Publique, F-75013, Paris, France; 2. INSERM, UMR\_S 1136, Institut Pierre Louis d'Epidémiologie et de Santé Publique, F-75013, Paris, France; 3. INSERM, UMR 1181 "Biostatistics, Biomathematics, Pharmacoepidemiology and Infectious Diseases" (B2PHI), F-75015 Paris, France; 4. Institut Pasteur, UMR 1181, B2PHI, F-75015 Paris, France; 5. Univ. Versailles St Quentin, UMR 1181, B2PHI, F-78180 Montigny-le Bretonneux, France; 6. Laboratoire MESuRS, Conservatoire National des Arts et Métiers, 75003, Paris, France; 7. INSERM, U1173, UFR Simone Veil, Versailles-Saint-Quentin University, 78180 Saint-Quentin en Yvelines, France; 8. AP-HP, Hôpital Raymond Poincaré, Service de Microbiologie, F-92380, Garches, France; 9. ENS de Lyon, Université de Lyon, Laboratoire de l'Informatique du Parallélisme (UMR CNRS 5668-ENS de Lyon-UCB Lyon 1), IXXI Rhône Alpes Complex Systems Institute, France; 10. INRIA-Institut National de Recherche en Informatique et en Automatique, France; 11. AP-HP, Hôpital Saint-Antoine, Département de Santé Publique, F-75571, Paris, France; 12. AP-HP, Raymond Poincaré Hospital, F-92380 Garches, France.

\*E.F., P.-Y.B., and D.G. equally contributed to this work.

†Members of the Individual-Based Investigation of Resistance Dissemination(i-Bird) Study Group are listed at the end of the text.

Received November 13, 2014; accepted March 26, 2015

© 2015 by The Society for Healthcare Epidemiology of America. All rights reserved. DOI: 10.1017/ice.2015.89

provide a good proxy for between-humans contacts and inform on interactions leading to transmission.<sup>13</sup>

Here, we report on a longitudinal, observational study, in which all participants were monitored for 4 months to continuously record CPIs and weekly detect *Staphylococcus aureus* carriage. The study took place in a long-term care facility (LTCF) in France. LTCFs, where patients usually stay for extended periods, present a large prevalence of carriage. As such, they have long been identified as a potential reservoir of antibiotic-resistant bacteria such as *S. aureus*.<sup>14–17</sup>

We conducted a nested case-control analysis to identify CPI-related risk factors associated with *S. aureus* transmission. Our objective was to identify CPI patterns that were associated with colonization. To that aim, we studied both local features of the CPI network (restricted to an individual of interest and his direct contacts) and global network measures. We also investigated whether the joint analysis of CPIs and bacterial carriage provided more information on the risk of colonization than mere CPI patterns.

## METHODS

### Study Organization and Setting

The I-Bird (Individual-Based Investigation of Resistance Dissemination) study took place from May 1 through November 1, 2009, in a 200-bed follow-up care and rehabilitation hospital in Berck-sur-Mer, France. During the study, 329 distinct patients stayed in the hospital and the total hospital staff comprised 261 healthcare workers (HCWs). They were all informed of the study and could refuse to participate or cease participation at any time. All participants (patients and HCWs) agreed to wear a small wireless sensor that recorded their CPIs and to undergo weekly nasal swabs to monitor *S. aureus* carriage. Usual hygiene procedures in France rely mostly on hand hygiene (alcohol-based hand rub, hand washing, and gloving for nursing cares) and do not implement routine screening for bacterial carriage. The complete i-Bird study protocol is detailed in the online supplementary text.

### *S. aureus* Carriage and Incident Colonization Episodes

*S. aureus* isolates were *spa*-typed<sup>18</sup> and tested for resistance to 20 antibiotics (see online supplementary text). Methicillin was used to distinguish between methicillin-sensitive (MSSA) and methicillin-resistant *S. aureus* (MRSA). Two strains were considered identical if they had the same *spa*-type and antibiotic resistance profile.

Incident colonization in a patient was defined when a strain was present that had not been carried the preceding week or, in case of multiple carriage, that had not been carried in the 2 preceding weeks.

### Recording CPIs Within the Hospital

Patients and staff wore a small wireless sensor at all times while in the hospital, recording the identity of others in a 1.5-m

range, front-facing for at least 30 s. The radio signal propagated in all directions but could sometimes be shielded by the body of participants. Patients wore the sensors on the wrist or ankle to limit positioning effects in the recordings. The sensor design is detailed in the online supplementary text. Recording was discontinued for 2 days on 3 occasions to replace sensor batteries.

There were 2,740,728 distinct CPIs recorded over the 4 months. These defined a dynamic (time-varying) network of interactions between participants, where individuals (nodes) were linked (edge) if a CPI had been recorded between them. This “full CPI network” was pruned of all CPIs lasting less than  $d$  min, varying  $d$  systematically between 0 and 15 min. In the text, we report the results for the network of all CPIs at least 5 min long.

The 1-hop CPI neighborhood of an individual included all his direct CPI contacts in the network; the 2-hop neighborhood included his 1-hop neighborhood and all the CPI contacts of his direct CPI contacts.

### Case and Controls Definition

The case-control study was nested in the longitudinal follow-up of the LTCF patients. A case was defined as a patient with incident colonization. Eight matched controls were selected as patients present in the LTCF at the date of the case's incident colonization but who were neither colonized with *S. aureus* at this date nor the week before.

### Statistical Analysis

We used conditional logistic regression to analyze factors associated with incident MRSA colonization. Our primary aim was the investigation of individual and CPI-related factors, including the following:

- Age, sex, Functional Independence Measure score,<sup>19</sup> history of MRSA carriage detected during the study,
- Number and occupation of CPI neighbors, duration of CPIs, density of links, centrality, number of neighbors-of-neighbors,
- Combination of network topology as above with carriage status: same MRSA strain or any MRSA strain carriage.

Time at risk was defined as the time between inclusion and date of incident colonization episodes (cases) or date of selection (controls). Functional Independence Measure scores were recoded as quartiles. For all network measures, the networks were aggregated over 1 week (the time interval between swabs) by pairs of participants. The betweenness centrality (the fraction of all shortest paths connecting pairs of participants passing through the person of interest) was computed using the whole network. We analyzed CPI durations directly (hours) and apportioned according to occupation of the contacts (patients or HCWs in contact with the case). As a sensitivity analysis, all analyses were repeated with various pruned CPI networks. The

Jaccard index,<sup>20</sup> a measure of the percentage of common edges in 2 networks, was used for comparisons.

## Ethics

Authorizations were obtained in accordance with French regulations regarding medical research and information processing. All French institutional review board-equivalent agencies accorded the i-Bird program official approval (CPP 08061; Afssaps 2008-A01284-51; CCTIRS 08.533; CNIL AT/YPA/SV/SN/GDP/AR091118 N°909036). Signed consent was not required by the French ethics committees to which the project was submitted.

## RESULTS

The median (range) age of patients was 58.5 (24.5–102.8) years, with a median (range) length of stay of 33 (1–158) days. The median (range) number of swabs collected in patients was 5 (1–16) and in HCWs was 8 (1–13). The full CPI network was dense, linking 20% of all possible pairs of individuals and clustered according to wards. Most CPIs were short (Table 1). As the CPI network was pruned of short CPIs, the Jaccard index decreased less quickly than the number of links, showing that most short CPIs doubled with a longer interaction with the same person. The CPI network based on CPIs longer than 5 min included a quarter of all CPIs (26%) but still had 38% edges in common with the full CPI network.

## Incident Colonization Episodes

Overall, 4,175 swabs were obtained from the participants. *S. aureus* prevalence of carriage was 37.1% (1,550 swabs in 363 participants) with few instances of multiple carriage (90 swabs from 65 participants showed 2 strains). Various strains ( $n = 114$ ) were isolated multiple times in different participants, indicating that transmission occurred. There were 237 incident colonization episodes in patients: 144 involved MRSA and 93 MSSA. Only MRSA episodes were investigated. After excluding episodes for which sensor data was not available at the time of sampling, 106

MRSA episodes were left for analysis. Eight controls were matched to each case unless too few were available (those cases were matched to 7 [ $n = 1$ ], 6 [ $n = 1$ ], 5 [ $n = 6$ ], 2 [ $n = 4$ ], and 1 [ $n = 2$ ] controls). Five cases were discarded because they could not be matched to any control. Fig. 1 illustrates the selection of cases and matching controls and the comparison of their neighborhoods.

## Individual Risk Factors for MRSA Colonization

Age or sex did not affect the risk of MRSA colonization (Table 2). Cases had been at risk for a longer time than controls (53.6 d vs 61 d,  $P = .002$ ), with a 5% (95% CI, 2%–9%) increase per additional week ( $P = .003$ ). Patients with lower functional ability scores were more prone to MRSA colonization. A history of previous MRSA carriage was more frequent in cases (OR, 24.53 [95% CI, 12.47–48.27],  $P < .0001$ ). Antibiotic exposure was associated with increased incidence (OR, 3.01 [95% CI, 1.63–5.59],  $P = .0005$ ).

## Network-Level Risk Factors for MRSA Colonization

Having more distinct neighbors during the preceding week was associated with an increased colonization risk (OR, 1.05 [95% CI, 1.01–1.09] for 1 additional neighbor,  $P = .009$ ; Table 3), more so with HCW neighbors than with patient neighbors. The risk was larger with neighbors in the same ward (OR, 1.09 [95% CI, 1.04–1.14],  $P = .0009$ ) and with night shift staff (OR, 1.29 [95% CI, 1.04–1.59],  $P = .019$ ).

Increased cumulated CPI duration with HCWs was associated with increased incidence (3% more per hour) as well as a higher share of one's CPIs spent with HCWs (OR, 1.15 [95% CI, 1.06–1.25] for a 10% increase,  $P = .001$ ). The betweenness centrality was not associated with the risk of colonization and neither was the density of the 1-hop neighborhood.

Interestingly, having CPI with contacts who themselves had many CPIs was associated with an increased colonization risk.

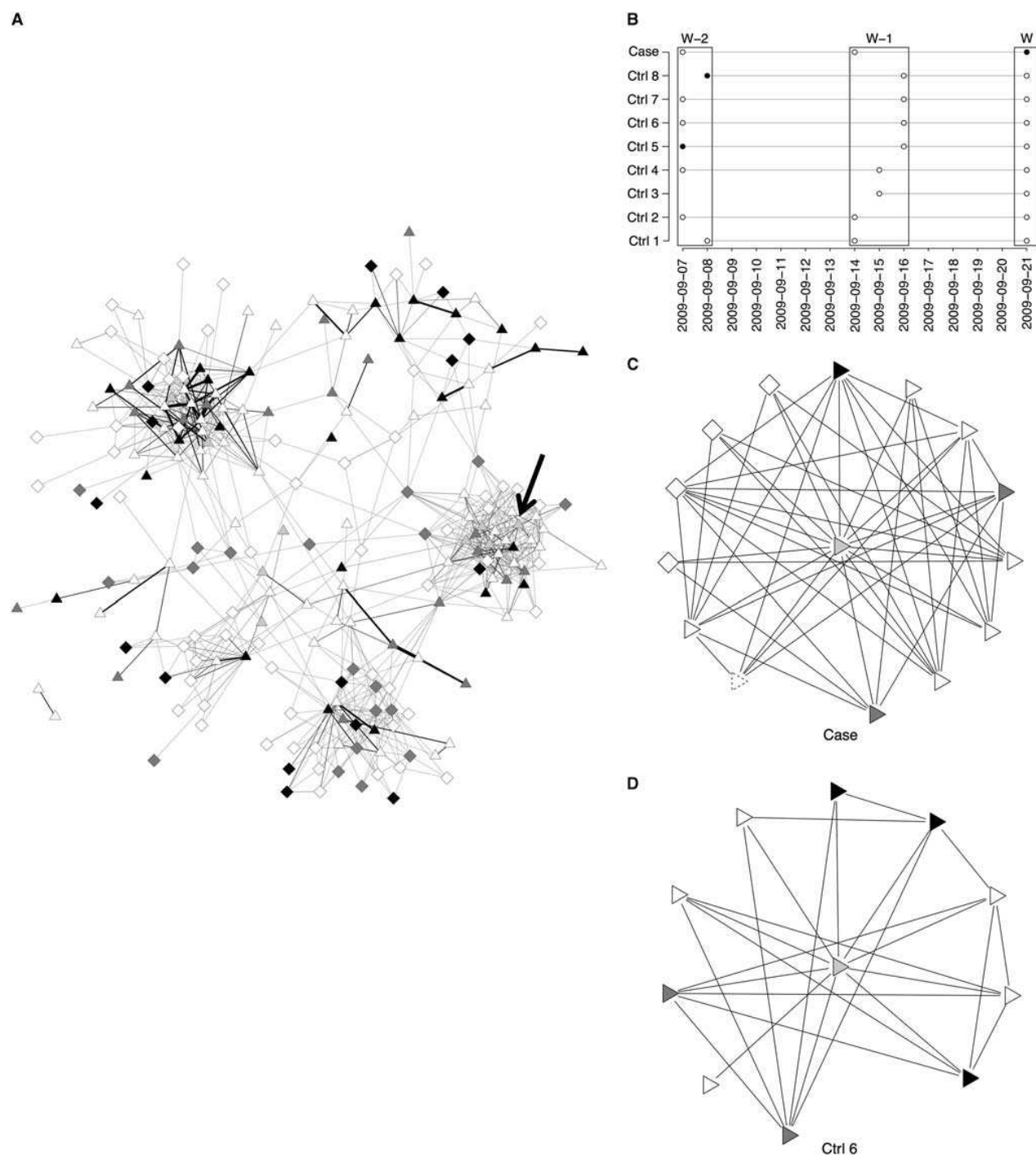
Associations were found at all levels of pruning of the full CPI network, with increasing ORs as shorter CPIs were discarded (Fig. 2).

TABLE 1. Characteristics of Complete and Thinned Close-Proximity Interaction (CPI) Networks

Network threshold, min	0	1	2	5	10	15
Network size, n	590	589	588	581	567	552
(patients; HCWs)	(329; 261)	(329; 260)	(329; 259)	(328; 253)	(322; 245)	(317; 235)
Jaccard index <sup>1</sup>	1.0	0.86	0.68	0.38	0.17	0.11
CPIs, n	85,025	62,993	46,459	22,347	11,315	8,382
(before aggregation)	(1,320,364)	(699,524)	(398,338)	(143,318)	(61,692)	(38,728)
CPIs/pers/day	10 (6–16)	7 (4–12)	6 (3–9)	3 (2–5)	2 (1–3)	2 (1–2)
Cumulative CPI duration, h	72,634	67,478	61,362	49,342	40,382	35,868
CPI duration/pers/day	8.6 (6.4–12.2)	8.1 (6.2–11.6)	7.7 (5.7–10.9)	7.5 (5.2–9.6)	7.8 (5.4–9.8)	8.2 (5.2–10.4)

NOTE. Values are median (interquartile range) unless stated otherwise. HCW, healthcare worker.

<sup>1</sup>The Jaccard index (a measure of similarity between networks) corresponds to the percentage of edges in common between 2 networks  $G_1$  and  $G_2$ . Here, we show  $\text{Jaccard}(G_d, G_0)$ , that is, the fraction of edges in common between a (pruned) network  $G_d$  and the full CPI network  $G_0$ .



**FIGURE 1. Case-control matching in the reference close-proximity interaction network.** **A**, An example close-proximity interaction (CPI) network in the week preceding a positive swab specimen in a case (identified by the arrow). Patients are shown as triangles and healthcare workers as diamonds and are colored according to their colonization status (white: noncolonized; dark gray: methicillin-sensitive *Staphylococcus aureus*; black: methicillin-resistant *S. aureus* [MRSA]). Two individuals are linked with an edge when a CPI exists between them. Thicker edges correspond to larger cumulated times of interaction. The case and matched controls are shown in light gray. **B**, Longitudinal swab specimen results in the case and matched controls at the date of case incidence (week W) and 2 preceding weeks (W-1 and W-2), colored according to colonization status (white: noncolonized; black: MRSA). **C**, Subnetwork centered on the case (from panel B) and restricted to his direct contacts (1-hop neighborhood). **D**, Subnetwork centered on a matched control (from panel B) and restricted to his direct contacts (1-hop neighborhood).

### Joint Use of Individual Status and Carriage Data

We analyzed jointly the characteristics of CPIs with carriage status for the risk of colonization (Table 4). The associations were generally stronger than above for CPIs with a neighbor carrying the same MRSA strain (OR, 2.13 [95% CI, 1.40–3.26]), and larger for patients' contacts than for HCWs' contacts. Such associations were also found when considering carriage of any MRSA strain, although smaller (OR, 1.4 [95% CI, 1.2–1.6]). Longer CPIs with colonized patients led to an increased risk (OR, 1.03 [95% CI, 1.01–1.05],  $P = .006$ ).

A strong association was found with the number of carriers at 2-hops (OR for same strain, 6.1; OR for any strain, 1.7). Connection to such carriers via an HCW was associated with a larger risk than via another patient.

Here again, keeping only longer CPIs in the network generally led to stronger associations (Fig. 2). In HCWs, the strongest association was found when considering the network of all CPIs longer than 1 min—that is, keeping a large portion of short CPIs.

TABLE 2. Individual Risk Factors of MRSA Colonization in Patients

Covariable	OR [95% CI]	P value
Male sex	1.18 [0.78–1.79]	.427
Age	1.00 [0.99–1.02]	.866
Time at risk (weeks)	1.05 [1.02–1.09]	.003
Antibiotic exposure	3.01 [1.63–5.59]	.0005
Past carriage (case)		
any MRSA strain	10.88 [6.74–15.57]	<.0001
same MRSA strain	24.53 [12.47–48.27]	<.0001
FIM Score		.006
Q1	2.90 [1.46–5.77]	
Q2	2.67 [1.35–5.30]	
Q3	1.81 [0.88–3.70]	
Q4	1.0	

NOTE. FIM, Functional Independence Measure; MRSA, methicillin-resistant *Staphylococcus aureus*; OR, odds ratio.

### Investigating Confusion Factors for Contacts and MRSA Incidence

Patients with higher dependence or a history of incident MRSA colonization had CPI with more neighbors in the preceding

TABLE 3. Network-Level Risk Factors of MRSA Colonization in Patients

Network characteristics	OR [95% CI]	P value
<b>Case/control</b>		
Betweenness centrality (%)	1.09 [0.94–1.27]	.231
Number of CPIs per day	1.54 [1.03–2.30]	.034
Number of distinct neighbors	1.05 [1.01–1.09]	.009
by occupation		
Patients	1.03 [0.97–1.09]	.259
HCWs	1.10 [1.04–1.17]	.001
by location		
Same ward	1.09 [1.04–1.14]	.0009
Other ward	1.01 [0.96–1.06]	.633
Cumulative CPI duration, h	1.00 [1.00–1.00]	.069
by occupation		
with patients	1.00 [1.00–1.00]	.106
with HCWs	1.03 [1.01–1.06]	.008
Time repartition (% with HCWs)	1.15 [1.06–1.25]	.001
<b>Contacts</b>		
Betweenness centrality (%)	0.95 [0.73–1.23]	.700
Number of CPIs per day	1.41 [1.00–1.99]	.052
Number of distinct neighbors	1.08 [1.02–1.15]	.012
by occupation		
Patients	1.05 [0.96–1.15]	.213
HCWs	1.15 [1.05–1.26]	.004
Cumulative CPI duration, h	1.01 [1.00–1.01]	.003
by occupation		
with patients	1.01 [1.00–1.01]	.005
with HCWs	1.10 [1.01–1.20]	.029
<b>Overall</b>		
Network density (%)	0.98 [0.88–1.09]	.681

NOTE. Depending on the characteristic, odds ratios (ORs) were computed for 1 more neighbor, 1 more hour cumulative close-proximity interaction (CPI) duration, or a 10% increase. HCW, healthcare worker; MRSA, methicillin-resistant *Staphylococcus aureus*.

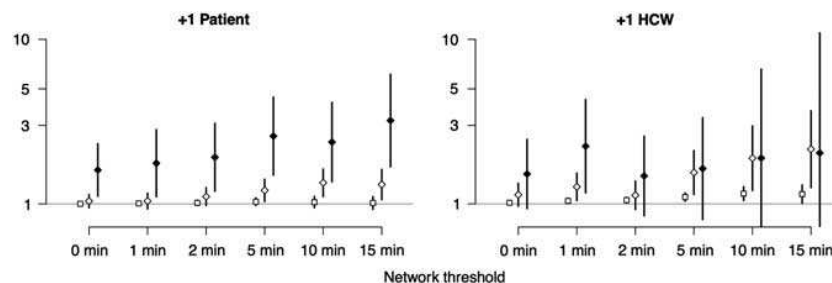


FIGURE 2. Close-proximity interaction (CPI) network topology and methicillin-resistant *Staphylococcus aureus* (MRSA) colonization. The odds ratios are for the number of contacts for incidence of MRSA colonization, according to occupation. White squares correspond to CPI network topology-only measures. Diamonds correspond to the joint analysis of carriage and CPI network topology when neighbors carry any MRSA strain (white) or the same MRSA strain as the case (black).

TABLE 4. Joint Microbiological and Network-Based Risk Factors of MRSA Colonization in Patients

Network characteristics	OR [95%CI]	P value
Distinct neighbors carrying		
Any MRSA strain	1.39 [1.21–1.60]	<.0001
Same MRSA strain	2.13 [1.40–3.26]	.0005
Distinct patient neighbors carrying		
Any MRSA strain	1.21 [1.03–1.41]	.017
Same MRSA strain	2.58 [1.49–4.46]	.0007
Distinct HCW neighbors carrying		
Any MRSA strain	1.55 [1.14–2.11]	.005
Same MRSA strain	1.64 [0.80–3.35]	.174
Distinct 2-hop neighbors carrying		
Any MRSA strain	1.66 [1.31–2.10]	<.0001
Same MRSA strain	6.05 [2.82–12.98]	<.0001
Any MRSA strain via patients	1.07 [0.81–1.40]	.649
Same MRSA strain via patients	4.30 [1.61–11.51]	.004
Any MRSA strain via HCWs	1.98 [1.50–2.61]	<.0001
Same MRSA strain via HCWs	19.77 [5.60–69.81]	<.0001
Cumulative CPI duration (h) with colonized neighbors		
Any MRSA strain	1.01 [1.00–1.02]	.0002
Patients	1.01 [1.00–1.02]	.0003
HCWs	1.56 [1.18–2.06]	.002
Same MRSA strain	1.03 [1.01–1.05]	.006
Patients	1.03 [1.01–1.05]	.006
HCWs	0.96 [0.28–3.23]	.942
Time repartition (% with carriers of any MRSA)	1.16 [1.09–1.23]	<.0001
with MRSA-carrying patients	1.13 [1.06–1.21]	.0002
with MRSA-carrying HCWs	2.00 [1.29–3.09]	.002

NOTE. Depending on the characteristic, odds ratios (ORs) were computed for 1 more neighbor, 1 more hour cumulative close-proximity interaction (CPI) duration, or a 10% increase. HCW, healthcare worker; MRSA, methicillin-resistant *Staphylococcus aureus*.

week (analysis of variance,  $P=.005$ ). These 2 characteristics were also associated with an increased colonization risk. After adjustment on the Functional Independence Measure score, the association of colonization with the number of distinct HCWs and longer CPIs with HCWs remained with similar odds ratios. Likewise, accounting for previous MRSA carriage did not change the associations. Cases still had had more CPIs during the week (OR, 1.02 [95% CI, 1.00–1.03] for 1 more neighbor;  $P=.02$ ).

## DISCUSSION

The characteristics of CPIs between individuals, especially between patients and HCWs, were associated with the risk of incident MRSA colonization in patients. Given our understanding of MRSA transmission,<sup>2</sup> these associations are expected and provide indication that CPIs capture contacts relevant to the spread of MRSA. However, summary characteristics of the overall network such as density did not

correlate with MRSA incidence because they remained approximately constant over time. Centrality was not associated with incidence either, although theoretical studies suggest a major role in transmission.<sup>4</sup>

Whether CPIs provide a good proxy for contacts leading to transmission is debated. Indeed, interactions with little risk of transmission (talking) and those at higher risk (nursing care) may lead to the same CPI records. However, this lack of specificity should lead to no association of CPIs with incidence because the same patterns would be found in case and in controls. Here, we found an association and it was the strongest for CPIs of longer duration. Yet, contacts alone cannot explain MRSA incidence unless contacts are carriers. The increased associations when considering CPIs with carriers of the same strain strengthened the case for using CPIs as a proxy for contacts leading to transmission.

Irrespective of the carried MRSA strain, CPI with a MRSA carrier was associated with increased incidence. It is known that MRSA carriage in HCWs is linked with poor adherence to infection control measures, exposure to highly prevalent patients, high work load, and other organizational factors.<sup>21</sup> HCWs colonized with a specific MRSA strain may be prone to be transient carriers and transmit other MRSA strains, because transient carriage of *S. aureus* is common among HCWs.<sup>22–24</sup> In our study, transient carriage was likely underdocumented, in part owing to the frequency of swabbing (once a week only) and also because hand carriage, for example, was not tested. Therefore, we investigated only incident episodes in patients.

The results support the role of HCWs as vectors of MRSA dissemination.<sup>5,6,25</sup> CPIs with HCWs were indeed associated with higher incidence than CPIs with other patients. Long contacts, common in nursing care, were the most associated with MRSA incidence, as often hypothesized.<sup>10,26,27</sup>

We tried to limit the effect of biases in our design and analysis. Whereas a few MRSA clones accounted for most episodes (MRSA overall prevalence was 16.6%, with >60% belonging to *spa*-types t777 or t008), MSSA carriage, although common, showed fewer identical strains carried by different individuals so that the study of transmission was less relevant (data not shown). Control periods were contemporaneous with case-defining periods to ensure equal *S. aureus* colonization pressure.<sup>28</sup> Classification bias (missed incident cases) may have occurred because the sensitivity of MRSA detection in swabs is not 100%. However, all results were unchanged in the analysis when limiting to the first incident episode in each case. No differential effect was expected in the CPI measurements because the carriage status of participants was not known at the date of CPI collection. An overall change in behavior, for example increased adherence to hygiene practices, may have occurred (Hawthorne effect<sup>29,30</sup>). Yet the protocol was little intrusive, with no external observer to remind of the ongoing study. Initial changes, if any, were likely to have waned over the 6-month-long data collection to yield CPIs illustrative of typical behaviors. Duration of hospital stay, functional status,



and previous MRSA carriage were associated with both the number of contacts and MRSA incidence. Adjusting on these variables in the analysis did not change the results concerning association with contacts, suggesting no confusion from these. HCW compliance with hygiene measures was not considered explicitly in our study. Hygiene compliance is more common during prolonged interactions<sup>31–34</sup> and yet this is where we found the strongest associations. This could indicate that compliance may alter the overall prevalence rather than the nature of interactions leading to transmission.

Patients from LTCFs may frequently travel to other hospitals, acting as a reservoir for MRSA introduction in other facilities.<sup>35</sup> In LTCFs, social interactions between patients occur more frequently than in acute care hospitals. We nevertheless found that CPIs with HCWs were associated with MRSA incidence, a result that is relevant in other settings.

We presented evidence that CPIs measured by electronic devices are associated with an increased risk of MRSA incidence. In our study, these devices were well accepted and worn for extended periods by all participants. The results suggest that routine CPI monitoring could be used to document in-hospital contact patterns and identify situations at increased risk. It could be a useful complement to microbiologic surveillance in identifying origins and destination of transmission.

#### Members of the i-Bird Study Group

The following people are members of the i-Bird Study Group: Anne Sophie Alvarez (AP-HP, Paris, France), Audrey Baraffe (AP-HP, Paris, France), Mariano Beiró (Universidad de Buenos Aires, Buenos Aires, Argentina), Inga Bertucci (AP-HP, Paris, France), Pierre-Yves Boëlle (Univ. Pierre et Marie Curie, Paris, France), Camille Cyncynatus (AbAg, Chilly-Mazarin, France), Florence Dannet (AP-HP, Paris, France), Marie Laure Delaby (AP-HP, Paris, France), Pierre Denys (AP-HP, Paris, France), Matthieu Domenech de Cellès (Univ. Pierre et Marie Curie, Paris, France), Eric Fleury (ENS Lyon, Lyon, France), Antoine Fraboulet (Insa, Lyon, France), Jean-Louis Gaillard (AbAg, Chilly-Mazarin, France), Didier Guillemot (Univ. Versailles Saint-Quentin, Versailles, France), Jean-Louis Herrmann (Univ. Versailles Saint-Quentin, Versailles, France), Boris Labrador (AP-HP, Paris, France), Jennifer Lasley (Inserm, Paris, France), Christine Lawrence (AP-HP, Paris, France), Judith Legrand (Univ. Paris Sud, Orsay, France), Odile Le Minor (Institut Pasteur, Paris, France), Caroline Ligier (Institut Pasteur, Paris, France), Lucie Martinet (Inria, Lyon, France), Karine Mignon (AbAg, Chilly-Mazarin, France), Lulla Opatowski (Univ. Versailles Saint-Quentin, Versailles, France), Catherine Sacleux (AP-HP, Paris, France), Jérôme Salomon (CNAM, Paris, France), Laura Temime (CNAM, Paris, France), Thomas Obadia (Univ. Pierre et Marie Curie, Paris, France), Marie Perard (AP-HP, Paris, France), Laure Petit (Institut Pasteur, Paris, France), Laetitia Remy (AP-HP, Paris, France), Anne Thiebaut (INSERM, Paris, France), Damien Thomas (AbAg,

Chilly-Mazarin, France), Philippe Tronchet (AP-HP, Paris, France), and Isabelle Villain (AP-HP, Paris, France).

#### ACKNOWLEDGMENTS

*Financial support.* European Commission under the Life Science Health Priority of the 6th Framework Program (MOSAR network contract LSHP-CT-2007-037941); French Government through the National Clinical Research Program and the Investissement d'Avenir program, Laboratoire d'Excellence "Integrative Biology of Emerging Infectious Diseases" (grant ANR-10-LABX-62-IBEID). The study sponsors had no role in study design, data collection, data analysis, data interpretation, or writing the report.

*Potential conflicts of interest.* All authors report no conflicts of interest relevant to this article.

Address correspondence to Thomas Obadia, INSERM U1136, 27 rue Chaligny, 75571 Paris CEDEX 12, France (thomas.obadia@upmc.fr).

#### SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit <http://dx.doi.org/10.1017/ice.2015.89>

#### REFERENCES

1. Pittet D, Allegranzi B, Storr J, Donaldson L. "Clean Care Is Safer Care": the Global Patient Safety Challenge 2005–2006. *Int J Infect Dis* 2006;10:419–424.
2. Pittet D, Allegranzi B, Sax H, et al. Evidence-based model for hand transmission during patient care and the role of improved practices. *Lancet Infect Dis* 2006;6:641–652.
3. Derde LPG, Cooper BS, Goossens H, et al. Interventions to reduce colonisation and transmission of antimicrobial-resistant bacteria in intensive care units: an interrupted time series study and cluster randomised trial. *Lancet Infect Dis* 2014;14:31–39.
4. Ueno T, Masuda N. Controlling nosocomial infection based on structure of hospital social networks. *J Theor Biol* 2008;254: 655–666.
5. Temime L, Opatowski L, Pannet Y, Brun-Buisson C, Boëlle PY, Guillemot D. Peripatetic health-care workers as potential super-spreaders. *Proc Natl Acad Sci* 2009;106:18420–18425.
6. Hornbeck T, Naylor D, Segre AM, Thomas G, Herman T, Polgreen PM. Using sensor networks to study the effect of peripatetic healthcare workers on the spread of hospital-associated infections. *J Infect Dis* 2012;206:1549–1557.
7. Cooper BS, Kyraios T, Batra R, Wyncoll D, Tosas O, Edgeworth JD. Quantifying type-specific reproduction numbers for nosocomial pathogens: evidence for heightened transmission of an Asian sequence type 239 MRSA clone. *PLOS Comput Biol* 2012; 8. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3325179/>. Accessed October 5, 2012.
8. Köser CU, Holden MTG, Ellington MJ, et al. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N Engl J Med* 2012;366:2267–2275.
9. Gardy JL, Johnston JC, Sui SJH, et al. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N Engl J Med* 2011;364:730–739.
10. Salathé M, Kazandjieva M, Lee JW, Levis P, Feldman MW, Jones JH. A high-resolution human contact network for

- infectious disease transmission. *Proc Natl Acad Sci* 2010;107:22020–22025.
11. Lucet J-C, Laouenan C, Chelius G, et al. Electronic sensors for assessing interactions between healthcare workers and patients under airborne precautions. *PLOS ONE* 2012;7:e37893.
  12. Smieszek T, Barclay VC, Seeni I, et al. How should social mixing be measured: comparing web-based survey and sensor-based methods. *BMC Infect Dis* 2014;14:136.
  13. Obadia T, Silhol R, Opatowski L, et al. Detailed contact data and the dissemination of *Staphylococcus aureus* in hospitals. *PLOS Comput Biol* 2015;11:e1004170.
  14. High KP, Bradley SF, Gravenstein S, et al. Clinical practice guideline for the evaluation of fever and infection in older adult residents of long-term care facilities: 2008 update by the Infectious Diseases Society of America. *Clin Infect Dis* 2009;48:149–171.
  15. Manzur A, Gudiol F. Methicillin-resistant *Staphylococcus aureus* in long-term-care facilities. *Clin Microbiol Infect* 2009;15:26–30.
  16. Bradley SF, Terpenning MS, Ramsey MA, et al. Methicillin-resistant *Staphylococcus aureus*: colonization and infection in a long-term care facility. *Ann Intern Med* 1991;115:417–422.
  17. Bilavsky E, Lerman Y, Rabinovich A, et al. Carriage of methicillin-resistant *Staphylococcus aureus* on admission to European rehabilitation centres—a prospective study. *Clin Microbiol Infect* 2012;18:E164–E169.
  18. Frénay HME, Bunschoten AE, Schouls LM, et al. Molecular typing of methicillin-resistant *Staphylococcus aureus* on the basis of protein A gene polymorphism. *Eur J Clin Microbiol Infect Dis* 1996;15:60–64.
  19. Keith RA, Granger CV, Hamilton BB, Sherwin FS. The functional independence measure: a new tool for rehabilitation. *Adv Clin Rehabil* 1987;1:6–18.
  20. Jaccard P. The distribution of the flora in the Alpine zone. *New Phytol* 1912;11:37–50.
  21. Albrich WC, Harbarth S. Health-care workers: source, vector, or victim of MRSA? *Lancet Infect Dis* 2008;8:289–301.
  22. Blok HEM, Troelstra A, Kamp-Hopmans TEM, et al. Role of healthcare workers in outbreaks of methicillin-resistant *Staphylococcus aureus*: a 10-year evaluation from a Dutch university hospital. *Infect Control Hosp Epidemiol* 2003;24:679–685.
  23. Ben-David D, Mermel LA, Parenteau S. Methicillin-resistant *Staphylococcus aureus* transmission: the possible importance of unrecognized health care worker carriage. *Am J Infect Control* 2008;36:93–97.
  24. Cookson B, Peters B, Webster M, Phillips I, Rahman M, Noble W. Staff carriage of epidemic methicillin-resistant *Staphylococcus aureus*. *J Clin Microbiol* 1989;27:1471–1476.
  25. McBryde ES, Pettitt AN, McElwain DLS. A stochastic mathematical model of methicillin resistant *Staphylococcus aureus* transmission in an intensive care unit: predicting the impact of interventions. *J Theor Biol* 2007;245:470–481.
  26. Meyers LA, Pourbohloul B, Newman MEJ, Skowronski DM, Brunham RC. Network theory and SARS: predicting outbreak diversity. *J Theor Biol* 2005;232:71–81.
  27. Newman MEJ. Spread of epidemic disease on networks. *Phys Rev E* 2002;66:016128.
  28. Harris AD, Samore MH, Lipsitch M, Kaye KS, Perencevich E, Carmeli Y. Control-group selection importance in studies of antimicrobial resistance: examples applied to *Pseudomonas aeruginosa*, enterococci, and *Escherichia coli*. *Clin Infect Dis* 2002;34:1558–1563.
  29. Pittet D, Simon A, Hugonnet S, Pessoa-Silva CL, Sauvan V, Perneger TV. Hand hygiene among physicians: performance, beliefs, and perceptions. *Ann Intern Med* 2004;141:1–8.
  30. Eckmanns T, Schwab F, Bessert J, et al. Hand rub consumption and hand hygiene compliance are not indicators of pathogen transmission in intensive care units. *J Hosp Infect* 2006;63:406–411.
  31. Boyce JM, Pittet D. Guideline for hand hygiene in health-care settings: recommendations of the Healthcare Infection Control Practices Advisory Committee and the HICPAC/SHEA/APIC/IDSA Hand Hygiene Task Force. *Infect Control Hosp Epidemiol* 2002;23:S3–S40.
  32. Eveillard M, Hitoto H, Raymond F, et al. Measurement and interpretation of hand hygiene compliance rates: importance of monitoring entire care episodes. *J Hosp Infect* 2009;72:211–217.
  33. Borg MA, Benbachir M, Cookson BD, et al. Self-protection as a driver for hand hygiene among healthcare workers. *Infect Control Hosp Epidemiol* 2009;30:578–580.
  34. Lau T, Tang G, Mak K, Leung G. Moment-specific compliance with hand hygiene. *Clin Teach* 2014;11:159–164.
  35. Lee BY, Bartsch SM, Wong KF, et al. The importance of nursing homes in the spread of methicillin-resistant *Staphylococcus aureus* (MRSA) among hospitals. *Med Care* 2013;51:205–215.

### **4.3 Commentaires et perspectives**

Dans cet article, le résultat majeur est que les données collectées dans l'étude I-Bird permettent de mettre en évidence que la structure locale du réseau des patients est associée à des différences significatives de risque de colonisation par SARM. Nous trouvons notamment des résultats cohérents avec la littérature, où le personnel de santé est désigné comme jouant un rôle crucial dans la dissémination notamment par portage transitoire. Nous discutons dans les paragraphes qui suivent de plusieurs points non-abordés dans l'article.

#### **4.3.1 Sur la représentativité du réseau mesuré**

##### ***4.3.1.1 Agrégation par semaine des réseaux égocentrés***

En réalisant l'analyse sur des réseaux égocentrés se limitant aux contacts directs, nous nous affranchissons du problème de cohérence temporelle détaillé au chapitre précédent. L'agrégation par paires des contacts ayant eu lieu dans ces sous-réseaux représente donc réellement l'hétérogénéité individuelle des interactions, à l'échelle de la semaine.

##### ***4.3.1.2 Effet Hawthorne***

L'effet Hawthorne consiste en une modification du comportement des personnes étudiées, lorsqu'elles se savent observées.[126] En milieu hospitalier, plusieurs études ont montré que le personnel de santé avait une meilleure adhérence aux mesures d'hygiène.[44, 120] Eckmanns et al. observent une amélioration de près de 50% lorsque le personnel de santé se sait observé. Comme évoqué dans l'article, l'étude I-Bird est peu intrusive, dans le sens où le seul « rappel » de son déroulement est hebdomadaire, lors de l'écouvillonnage nasal. De plus, au regard de la durée de l'étude (6 mois), l'hypothèse qu'un tel effet, s'il a existé, s'est atténué avec le temps, semble raisonnable. Enfin, le fait que les deux premiers mois de collecte de données (mai et juin 2009) n'aient pas été pris en compte dans l'étude renforce encore cette hypothèse.

#### **4.3.2 Sur la définition des cas**

Le suivi longitudinal du portage permet de détecter l'acquisition d'une nouvelle souche chez un participant, ainsi que son éventuelle décolonisation. Un même individu peut donc générer plusieurs épisodes d'incidence au cours de son séjour, soit par l'acquisition

de plusieurs souches distinctes, soit par un phénomène de colonisation/décolonisation/recolonisation. En effet, il n'existe pas de vrai consensus sur la durée d'une colonisation à SARM : dans une revue systématique par Shenoy et al., le taux de décolonisation varie de 12% à 79%, pour des durées allant d'une à 208 semaines après la date de colonisation.[131] La décolonisation est par ailleurs associée à l'autonomie des patients.[100] Nous autorisons donc, sous certaines conditions détaillées dans l'article, une nouvelle colonisation chez un patient par une souche qu'il a déjà portée dans le passé, si elle n'a pas été isolée dans le prélèvement de la semaine précédente (et dans les deux derniers prélèvements en cas de portage de plusieurs souches, pour éviter des phénomènes de compétition et d'inhibition de croissance par une souche concurrente).

Dans l'article, nous présentons les résultats sur des épisodes de colonisation répondant à cette définition. La sensibilité imparfaite de la méthode d'échantillonnage (estimée entre 61.5% et 80%) fait que chaque prélèvement négatif est susceptible d'être un faux-négatif avec une probabilité non-négligeable (entre 20% et 38.5%). S'il est impossible d'identifier rétrospectivement ces prélèvements, une hypothèse raisonnable est que les prélèvements de type +/-/+ (pour la même souche) sont constitués pour majorité de faux négatifs. Nous avons réalisé la même étude cas-témoins en utilisant une version lissée des données de portage, en transformant les séquences +/-/+ en +/+/+ chez les patients. Ce jeu de donnée lissée conduit à l'identification de 127 épisodes d'incidence, dont 93 peuvent être inclus dans l'analyse (contre 144 et 101 précédemment). Les conclusions concernant les facteurs de risque sont inchangées. Les résultats de cette analyse, non mentionnée dans l'article, sont proposés aux Annexe 1. Dans l'article, nous rapportons une analyse de sensibilité similaire, mais cette fois limitée au premier épisode de colonisation chez chaque patient. De nouveau, les associations restent semblables.

#### **4.3.3 Sur les facteurs de risque identifiés**

##### ***4.3.3.1 Peu de significativité pour les colonisations par *S. aureus* sensible à la méticilline***

La même analyse cas-témoins a été menée sur les épisodes de colonisation à SASM. Dans ce volet de l'étude, le réseau des contacts ne permet pas d'identifier des facteurs de risque. Les résultats sont fournis en Annexe 2. Ce résultat n'est pas inattendu, puisque la diversité des souches SASM isolées était plus grande, et les SASM se trouvent

habituellement dans la communauté : l'acquisition de SARM par des patients peut provenir de l'extérieur (visites, sorties...), notamment dans cet hôpital où la plupart des patients sont admis pour une période longue, et sortent de l'hôpital s'ils le peuvent.

#### **4.3.3.2 Facteurs de risque « réseau »**

Le premier constat que nous pouvons faire est que la structure locale du réseau de contacts informe sur le risque de colonisation par SARM. Nous remarquons en effet que ce risque augmente légèrement avec le nombre de voisins (OR = 1.035 [1.002 ; 1.069] pour un voisin supplémentaire), et que cette augmentation est causée par des contacts avec de nouveaux personnels de santé plutôt que des patients (OR = 1.073 [1.016 ; 1.133] pour un personnel de santé supplémentaire vs. OR = 1.025 [0.975 ; 1.077] pour un patient supplémentaire). La même tendance est observée pour les durées de contacts, où les contacts plus longs avec des personnels de santé sont associés à un risque plus élevé de colonisation (OR = 1.034 [1.008 ; 1.060] pour une heure supplémentaire). Ces résultats sont cohérents avec le rôle supposé des personnels de santé, qui peuvent agir comme porteurs transitoires entre deux patients. Un découpage plus fin du personnel de santé (personnel de jour du service, ergothérapeute, kinésithérapeute, personnel de nuit, administratif) fait apparaître un risque accru uniquement avec la catégorie « personnel de nuit » (OR = 1.289 [1.052 ; 1.593] pour un personnel de nuit supplémentaire). Ceci est à mettre en lien avec les résultats de Temime et al. et Hornbeck et al., [76, 137] et tend à confirmer que certains profils de personnel de santé provoquent des hausses d'incidence significatives.

#### **4.3.3.3 Facteurs de risque « réseau et microbiologie »**

Comme on pourrait l'attendre, l'étude conjointe des contacts et de leur dernier portage connu fournit des associations plus fortes pour les contacts avec des personnes colonisées. Un résultat intéressant est toutefois que ces associations perdent en significativité lorsque l'on considère les contacts avec des personnels de santé porteurs de la même souche que celle du patient incident (pour exemple, voir Figure 2 dans l'article).

Nous expliquons cette perte de significativité par une perte de puissance statistique, causée par la granularité des données de portage qui interdit la détection du portage transitoire. Toutefois, le portage d'un SARM (indépendamment de la souche précise) permet de récupérer une partie du signal et produit des associations significatives,

même pour les contacts avec des personnels de santé : le portage d'un SARM peut permettre de détecter des personnels de santé colonisés, et peut-être plus à risque de transmettre d'autres souches lors d'événements de colonisation transitoires dus à une mauvaise adhérence aux mesures d'hygiène.

\* \* \*

Cette étude permet donc, en utilisant des données descriptives relativement simples, d'identifier des sources de risque accru pour les patients. Nous identifions des facteurs de risques mesurables directement par « l'activité » de chaque patient sur le réseau. Un point notable est que les mesures tenant compte de l'ensemble du réseau des contacts ne sont pas à même de capturer ces risques. Ceci est à mettre en parallèle avec l'effet « petit-monde » déjà évoqué aux chapitres précédents : les individus se retrouvent très rapidement connectés à tous les autres, et les propriétés du réseau sont constantes d'un jour à l'autre. Dans le chapitre suivant, nous allons aborder plusieurs pistes qui peuvent permettre de mettre en place des stratégies de contrôle tenant compte du réseau observé, afin de limiter la dissémination des SARM dans la structure étudiée.

## **5 Chapitre 5 : Dynamiques de formation et de dissolution de liens et considérations épidémiologiques dans un réseau à topologie variable**

Nous avons vu au fil des chapitres précédents que les données de l'étude I-Bird permettent de capturer des contacts qui modulent les risques de colonisation par *S. aureus* de patients hospitalisés. Nous allons maintenant évoquer les perspectives d'exploitation possibles de ces données. Dans un premier temps, nous étudions plusieurs méthodes qui peuvent permettre la reconstruction de données manquantes. Ensuite, nous analyserons l'organisation temporelle des interactions à une échelle plus fine que celle utilisée jusqu'alors. En discrétisant les données avec une résolution plus élevée, nous proposerons enfin plusieurs pistes pour permettre de quantifier des seuils épidémiques qui tiennent compte de la dynamique du réseau mesuré. Ces méthodes formeront une base de réflexion pour la mise au point de nouvelles mesures de contrôle pour limiter la dissémination des souches antibiorésistantes, tenant compte de l'hétérogénéité des comportements des participants. Dans ce chapitre, nous proposons plusieurs méthodes permettant de répondre aux problématiques soulevées.

### **5.1 Données manquantes et méthodologie d'échantillonnage du réseau**

On peut remarquer, sur la Figure 18 (Chapitre 2), que le nombre de nœuds dans le réseau est maximal juste après un changement de batteries, et diminue ensuite jusqu'au changement suivant. Le réseau mesuré lors de l'étude I-Bird est en réalité incomplet : les capteurs dont la batterie se vide avant la date de changement prévue ne sont pas remplacés en avance. Ces données manquantes ne doivent toutefois pas introduire de biais systématique par rapport aux données de colonisation, puisque celles-ci n'étaient pas communiquées aux participants. Les résultats évoqués dans les chapitres précédents, ainsi que dans les 2 articles, ne doivent donc pas être affectés puisque les cas et les témoins ne sont pas affectés de façon différentielle par ce phénomène.

Il n'en reste pas moins que nous n'observons qu'un sous-ensemble de tous les contacts qui ont vraiment eu lieu dans I-Bird. Cela nous amène à discuter le cadre dans lequel sont réalisés ces enregistrements.

### 5.1.1 Méthode d'échantillonnage des contacts

Dans I-Bird, chaque participant dispose de son propre capteur, identifié par un numéro unique. Ce capteur fonctionne par cycles de 30 s, au cours desquels se déroulent deux périodes (émission de l'identifiant/réception des identifiants émis par les voisins). Les données de chaque capteur fournissent donc un réseau égocentrique dirigé, dont la source est le porteur, et la cible est la personne contactée.

### 5.1.2 Notations

Nous reprenons dans la suite les notations de Handcock & Gile,[68] basées sur les travaux d'autres auteurs.[138, 139] Soit  $\mathcal{Y}$  l'univers des réseaux  $Y$  possibles pour l'ensemble  $V$  de  $n$  nœuds, et  $y$  la réalisation de  $Y$ ;  $D$  la matrice binaire de dimensions  $n * n$  (dont l'univers est  $\mathcal{D}$ ) qui indique si une paire d'individus a été échantillonnée (*i.e.*  $D_{ij} = 1$  si la paire orientée  $i \rightarrow j$  a été échantillonnée, 0 sinon). La distribution de  $D$  est supposée dépendre d'un paramètre  $\psi$ . Dans I-Bird,  $\psi$  correspondrait à l'état de la batterie des capteurs.

$P(D = d \mid Y = y, \psi)$  correspond à la probabilité de sélectionner l'échantillon  $d$  sachant un réseau (complet)  $y$  et  $\psi$ . Le réseau  $Y$  est constitué de sa fraction observée, et de sa fraction non-observée, *i.e.*  $Y = Y_{obs} + Y_{mis}$ .

Dans I-Bird, l'échantillonnage est conventionnel (au sens de Handcock & Gile) dans la mesure où il n'utilise pas d'informations sur les nœuds pour réorienter l'échantillonnage au cours de l'étude.[68] Les auteurs définissent par ailleurs ce type d'échantillonnage comme un cas particulier de design adaptatif, dans lequel l'échantillonnage peut être réajusté mais uniquement en fonction des données observées (*i.e.*  $P(D = d \mid Y, \psi) = P(D = d \mid Y_{obs}, \psi)$ ). Dans ce type d'échantillonnage, la distribution de  $D$  est uniquement conditionnelle à la partie observée du réseau.

### 5.1.3 Inférence pour un réseau résultant d'un modèle paramétrique

Lorsque le réseau  $Y$  est la réalisation d'un processus stochastique gouverné par un vecteur de paramètre  $\eta$ , la vraisemblance des paramètres  $\psi$  et  $\eta$  s'écrit :

$$\mathcal{L}(\eta, \psi \mid Y_{obs} = y_{obs}, D = d_{obs}) = P(Y_{obs} = y_{obs}, D = d_{obs} \mid \eta, \psi)$$

Quand  $\eta$  et  $\psi$  sont des paramètres indépendants, cette vraisemblance s'écrit comme le produit des vraisemblances marginales de chaque paramètre, et donc :

$$\mathcal{L}(\eta \mid Y_{obs} = y_{obs}) \propto P_{\eta}(Y_{obs} = y_{obs}) = \sum_{v \in Y_{mis}} P_{\eta}(Y = y_{obs} + v)$$



En maximisant cette vraisemblance, il est possible de trouver la configuration de  $Y_{mis}$  la plus plausible selon le modèle paramétrique postulé pour  $Y$ , car elle est définie à une constante multiplicative près (la vraisemblance de  $\psi$ ).

Dans notre cas, le processus d'échantillonnage (*i.e.* d'observation) est lié au fonctionnement du capteur, qui est *a priori* indépendant de la réalisation d'un contact entre deux personnes. Nous sommes donc dans le cas de données « Missing Completely At Random », en ce qui concerne le réseau des contacts.

## 5.2 Pistes pour l'imputation des données manquantes

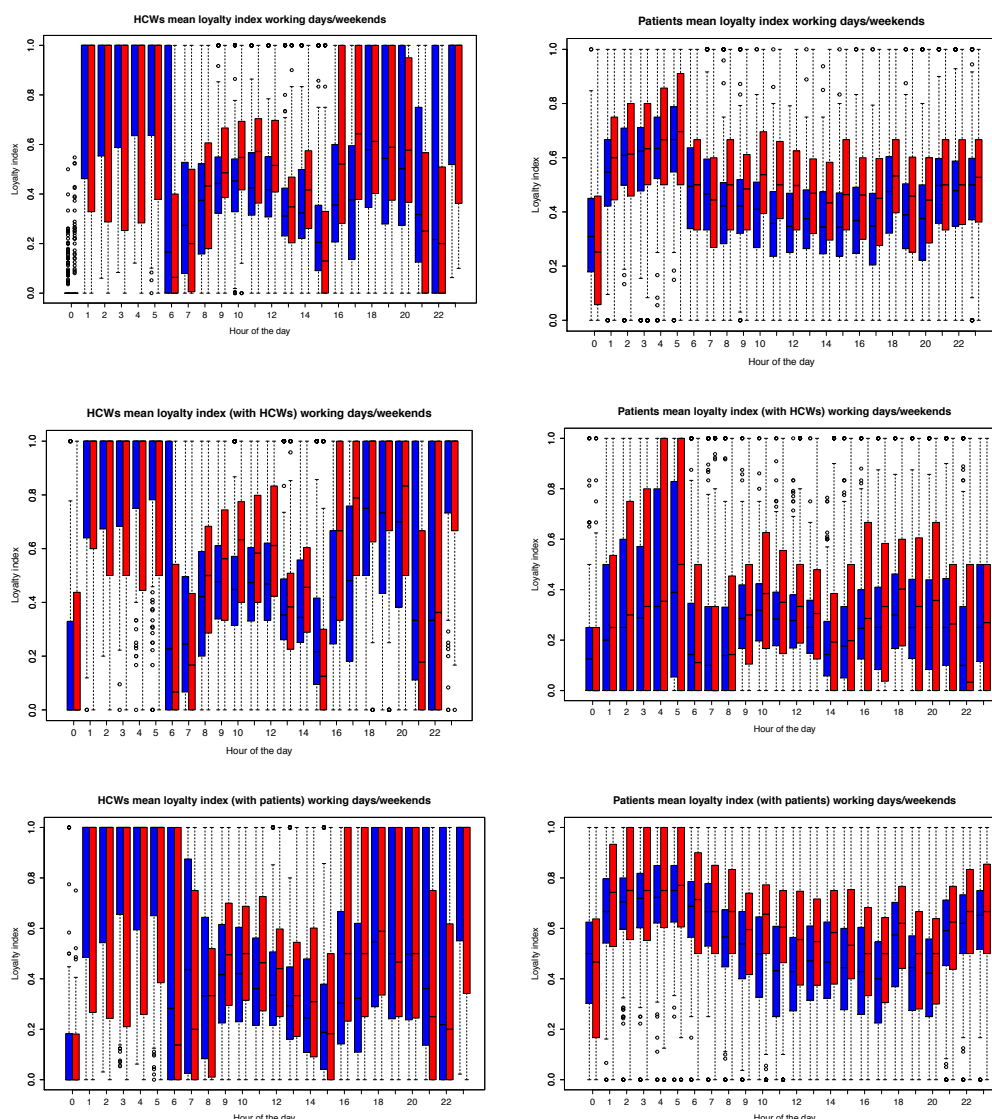
Dans la section précédente, nous avons introduit une formulation de la vraisemblance d'un paramètre lorsque le réseau est la réalisation d'un processus stochastique. Une approche descriptive est également envisageable. Nous proposons dans les paragraphes qui suivent une description de ces deux approches.

### 5.2.1 Imputation par la description du voisinage (approche empirique)

Nous décrivons partiellement dans le S2 Text de l'article présenté au paragraphe 3.3.2 que le voisinage direct d'un patient se construit très rapidement dès son inclusion. Quelques jours de données seulement permettent de déterminer de façon assez précise les voisins qu'il aura tout au long de son séjour dans l'hôpital. Les données de présence dans l'hôpital permettent également de déterminer si un individu devrait avoir des contacts un jour donné, et donc si son capteur est défaillant ou non. Nous pouvons alors échantillonner parmi ses anciens voisins connus pour lui rajouter des liens. Cet échantillonnage peut tenir compte de paramètres individuels comme le service dans lequel il est (pour respecter le clustering) et sa « loyauté », exprimée en fonction de sa propension à conserver des liens avec les mêmes individus d'un pas de temps à l'autre (par exemple, d'un jour à l'autre, ou d'une heure à l'autre). La loyauté entre deux pas de temps est définie par l'indice de Jaccard (avec  $G_1$  et  $G_2$  les graphes égocentriques du voisinage 1-hop d'un patient aux pas de temps  $t_1$  et  $t_2$ ):

$$Jaccard(G_1, G_2) = \frac{G_1 \cap G_2}{G_1 \cup G_2}$$

A titre d'exemple, la Figure 24 présente ces mesures pour un pas de temps d'une heure, avec distinction entre jours ouvrés et week-end.



**Figure 24. Évolution de l'indice de loyauté des participants entre deux tranches horaires successives, au cours de la journée. (Gauche) Indice de loyauté pour les personnels de santé. (Droite) Indice de loyauté pour les patients. Chaque encart représente l'indice de loyauté entre deux heures consécutives, avec l'ensemble des voisins (**haut**), les voisins patients (**milieu**) et les voisins personnels de santé (**bas**). Les couleurs correspondent aux jours de la semaine (bleu : jour ouvré ; rouge : week-end).**

Nous pouvons observer des schémas réguliers, qui correspondent (notamment pour le personnel de santé) aux changements des équipes, à horaire fixe (vers 6h, 14h et 22h).

Cette approche descriptive a été abordée récemment par Valdano et al. pour prédire le risque épidémique sur un réseau inconnu à partir d'anciennes observations.[143] En particulier, les auteurs montrent que le risque épidémique est plus élevé chez les individus ayant un indice de loyauté plus élevé, sur la base de deux réseaux de contacts différents (un sur des déplacements de bovins en Italie, et un autre sur des données de contacts sexuels dans un réseau de prostitution).

## 5.2.2 Imputation par modélisation de la formation des liens

L'approche proposée dans la section précédente repose sur une description correcte des liens déjà créés. Même si elle permet de reconstruire certaines arêtes manquantes dans le réseau, elle n'assure pas de capter tout le voisinage réellement rencontré. Une alternative est de modéliser les processus qui amènent à l'établissement d'un lien entre deux individus, pour simuler des réseaux similaires à ceux d'I-Bird, enrichis des nœuds pour lesquels les données sont absentes. Pour cela, nous proposons l'utilisation des Exponential Random Graph Models (ERGM).

### 5.2.2.1 Les ERGM (*Exponential Random Graph Models*)

Les ERGM cherchent à modéliser les processus sociaux qui entraînent à la formation d'un lien dans un réseau de contacts, à l'échelle locale.[56, 62, 63, 116, 122, 123, 134, 151]

#### 5.2.2.1.1 Écriture du modèle paramétrique pour $Y$

Le réseau  $Y$  est considéré comme une variable aléatoire, dont la réalisation dépend d'un  $p$ -vecteur de statistiques  $g(y)$  (description au paragraphe 5.2.2.1.2 ci-après) et d'un vecteur de paramètres  $\eta$ . Le modèle s'écrit :

$$P_{\eta}(Y = y) = \exp(\eta * g(y) - \kappa(\eta))$$

où  $\exp(\kappa(\eta)) = \sum_{u \in \mathcal{Y}} \exp(\eta * g(u))$  est une constante de normalisation. En pratique, cette constante est difficile à calculer étant donné le grand nombre de configurations de  $Y$  contenues par  $\mathcal{Y}$ . Des approches de calculs intensifs des méthodes de Monte Carlo par Chaînes de Markov (MCMC) sont utilisées pour l'approcher.[69, 78, 134]

#### 5.2.2.1.2 Description des statistiques contenues dans $g(y)$

Les termes permettant de paramétrer ce modèle sont documentés de façon détaillée par Morris et al.[107] Ce sont principalement des statistiques ayant trait à la topologie du réseau, ou à des attributs nodaux. Les termes de topologie sont par exemple la densité, le nombre d'arêtes, le nombre de triangles. Les termes liés aux attributs nodaux comptent le nombre de fois où deux individus partageant un attribut (qualitatif ou quantitatif) ont un lien les reliant. Ces attributs nodaux permettent notamment de capturer l'assortativité : âge, sexe, service d'hospitalisation, statut (patient ou personnel de santé)... Avec les données d'I-Bird, le terme « `nodematch(service)` » correspond au

nombre de paires de nœuds ayant le même attribut « service », c'est à dire le nombre de paires de liens qui impliquent deux nœuds d'un même service.

Les coefficients de  $\eta$  estimés par le modèle permettent de calculer le logarithme de la cote de l'événement « un lien se crée entre  $i$  et  $j$  » en fonction des changements que cet événement implique dans le vecteur  $g(y)$ . La probabilité que ces deux individus forment un lien est alors simple à calculer.

### 5.2.2.2 Imputation des liens manquants par des approches de calcul intensif

Nous pouvons continuer le raisonnement entamé à la section 5.1.3 en utilisant le contexte des ERGM. D'après Handcock & Gile, la distribution de  $Y$  conditionnelle à  $Y_{obs}$  se résume à celle de  $Y_{mis}$  [68] :

$$P_{\eta}(Y = Y_{obs} + v | Y_{obs}) = P_{\eta}(Y_{mis} = v | Y_{obs} = y_{obs}) = \exp(\eta * g(v + y_{obs}) - \kappa(\eta | y_{obs}))$$

où  $\kappa(\eta | y_{obs})$  est la constante de normalisation associée à l'espace des graphes restreint à ceux contenant  $y_{obs}$ . En écrivant la distribution de  $Y$  comme un produit de probabilités conditionnelles :

$$P_{\eta}(Y) = P_{\eta}(Y_{mis} | Y_{obs}) * P_{\eta}(Y_{obs})$$

la vraisemblance du vecteur de paramètres  $\eta$ ,  $\mathcal{L}(\eta | Y_{obs} = y_{obs})$ , devient proportionnelle au rapport des deux constantes de normalisation :

$$\mathcal{L}(\eta | Y_{obs} = y_{obs}) \propto \exp(\kappa(\eta | y_{obs}) - \kappa(\eta))$$

Ici, le vecteur  $\eta$  peut être estimé par deux chaînes MCMC : l'une sur le réseau conditionnel à  $Y_{obs}$  et la seconde sur le réseau  $Y$ . [68]

### 5.2.3 Exemple

Nous présentons ici les résultats de plusieurs modèles ERGM appliqués à une journée de contacts. Après avoir ajusté chaque modèle aux données de contacts du jour, 100 réseaux issus de modèle sont simulés, avec pour contrainte de conserver les arêtes observées dans le réseau d'origine. Les arêtes manquantes sont imputées, et nous pouvons comparer l'écart entre plusieurs statistiques descriptives du réseau initial et celles obtenues en moyennant les résultats des 100 simulations. Nous envisageons ici 5 modèles de complexité croissante, dont les résultats sont proposés dans la Table 1.

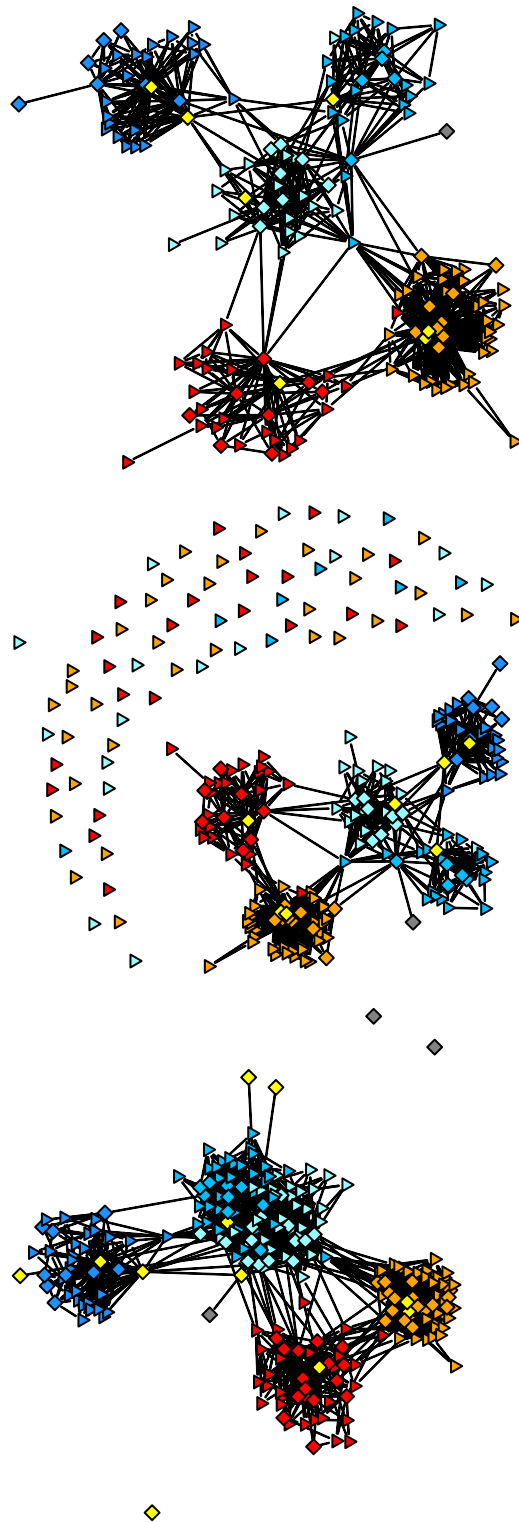
Le modèle M1 génère un graphe aléatoire d'Erdős-Rényi, et n'utilise donc qu'un seul terme (le nombre d'arêtes). Les modèles suivants sont obtenus par l'ajout successif :

- d'un terme d'homophilie sur le service de rattachement (M2),
- d'un terme d'homophilie sur le bâtiment de rattachement (M3),
- d'un terme tenant compte de l'activité (en nombre de liens) de chaque service, *i.e.* le nombre de fois ou chaque service est impliqué dans un lien (M4),
- d'un terme tenant compte du score d'indépendance fonctionnelle (M5).

Network	Target statistics							BIC
	Edges	Density	Triangles	Node mix			Node match	
				PA-PA (%)	PA-HCW (%)	HCW-HCW (%)	Ward	
Original	864	7.2%	1737	204 (23.6%)	553 (64%)	107 (12.4%)	661	NA
M1	1941	7.2%	770	866 (44.6%)	865 (44.6%)	210 (10.8%)	346	6210
M2	1485	5.4%	1733	597 (40.2%)	732 (49.3%)	156 (10.5%)	1178	3966
M3	1491	5.5%	1660	590 (39.6%)	753 (50.5%)	148 (9.9%)	1189	3935
M4	1520	5.6%	1927	627 (41.2%)	735 (48.4%)	158 (10.4%)	1161	3798
M5	1515	5.6%	1971	543 (35.8%)	769 (50.8%)	203 (13.4%)	1152	3679

**Table 1. Comparaison des données originales à 5 ERGM pour l'imputation des liens manquants sur une journée de contacts observés dans l'étude I-Bird.** Les statistiques des données originales ne tiennent pas compte des nœuds supposés manquants, à l'inverse de celles rapportées pour les modèles M1 à M5. Les pourcentages dans la section « node mix » sont relatifs au nombre total d'arêtes (« edges »).

Dans tous les cas, les réseaux générés ont une densité comparable à celle du réseau observé. Les modèles introduisant de l'assortativité par service rendent bien compte du clustering qui s'y produit (terme « node match » élevé par rapport au nombre global d'arête « edges »). En utilisant le modèle 5 (qui semble le plus performant, ayant un critère d'information bayésien (BIC) le plus faible), la Figure 25 présente le processus d'ajout des liens manquants. Ce modèle pourrait probablement encore être affiné, mais sa simplicité permet de recréer des réseaux aux propriétés similaires à celles des données observées.



**Figure 25. Étapes successives pour l'imputation des liens manquants dans le réseau I-Bird.** Nous utilisons les données observées au cours d'une journée. Les nœuds manquants sont déduits des emplois du temps du personnel et des données d'admission des patients. **(Haut)** Réseau observé. **(Milieu)** Réseau observé avec ajout des nœuds supposés manquants. **(Bas)** Réseau observé complété par les liens imputés par le modèle M5. Dans les trois schémas, les couleurs représentent le service de rattachement ou la catégorie de personnel de santé (orange/rouge : W1 et W2 ; 3 nuances de bleu : W3 à W5 ; jaune : personnel de nuit ; gris : administratif/autre).

### 5.2.4 Separable Temporal ERGM : extension aux réseaux dynamiques

Enfin, une extension du concept des ERGM a été proposée par Krivitsky & Handcock.[92] Les ST-ERGM proposent d'ajuster deux modèles à un réseau temporel : un pour la formation des liens, et un pour la dissolution. Toutefois, les calculs requis pour des réseaux comme ceux observés dans I-Bird deviennent rapidement très coûteux en temps. Il reste envisageable de modéliser la dynamique des liens en restreignant les données à une période fine, constituée de peu de pas de temps.

## 5.3 Perspectives épidémiologique de prédiction du risque épidémique

Dans le Chapitre 1, nous avons évoqué l'intérêt des réseaux de contacts pour le calcul de seuils épidémiques tenant compte de l'hétérogénéité individuelle. De récents travaux par Valdano et al. ont étendu l'approche markovienne du modèle S-I-S (présentée en section 1.5.4) en autorisant une variation de la matrice d'adjacence en fonction du temps.[142]

### 5.3.1 Extension du modèle S-I-S sur un substrat dynamique

En posant  $A^{(t)}$  la matrice d'adjacence pour le réseau observé au temps  $t$ , la probabilité de devenir infecté au temps  $t+1$ , pour l'individu  $i$  (notée  $P(i, t+1)$ ), dépend uniquement de la configuration précédente du réseau :

$$P(i, t + 1) = 1 - [1 - (1 - \mu) * P(i, t)] * \prod_{j \neq i} [1 - \lambda A_{ji}^{(t)} P(j, t)]$$

où  $\lambda$  et  $\mu$  sont les paramètres du modèle S-I-S. Les auteurs proposent une nouvelle forme de stockage du réseau, afin de conserver la corrélation temporelle des liens. Ils proposent l'écriture couplée de la dynamique S-I-S sur le substrat variable  $A^{(t)}$  comme une nouvelle matrice de dimensions  $n * T_{max}$  :

$$M = \begin{pmatrix} 0 & 1 - \mu + \lambda A^{(1)} & 0 & \dots & 0 \\ 0 & 0 & 1 - \mu + \lambda A^{(2)} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 - \mu + \lambda A^{(T_{max}-1)} \\ 1 - \mu + \lambda A^{(T_{max})} & 0 & 0 & \dots & 0 \end{pmatrix}$$

Avec cette notation multicouche, chaque nœud est relié à sa future image (il reste infecté *via* ce lien avec une probabilité  $1-\mu$ ) et à ses voisins actuels (chacun pouvant l'infecter avec une probabilité  $\lambda$ ). Cette nouvelle configuration du réseau permet de donner une dimension statique au problème temporel : le processus de Markov est décrit par une trajectoire dans un espace de dimensions  $n * T_{max}$ , et la probabilité d'infection, au cours

d'une période  $[\tau * T_{max}; (\tau + 1) * T_{max}]$ , pour un nœud  $\alpha$  (le nœud  $i$  au temps  $t$ , identifié dans ce nouvel espace de dimension  $n * T_{max}$ ), devient :

$$\widehat{p}_\alpha(\tau) = 1 - \prod_{\beta \neq \alpha} 1 - M_{\beta\alpha} \widehat{p}_\beta(\tau - 1)$$

dont la solution asymptotique vérifie :

$$\widehat{p}_\alpha = 1 - \prod_{\beta \neq \alpha} 1 - M_{\beta\alpha} \widehat{p}_\beta$$

Dans ces formulations,  $\alpha$  et  $\beta$  représentent des nœuds du réseau multicouches, donc la combinaison d'un identifiant ( $i$  dans le processus Markovien non dynamique) et d'un pas de temps ( $t$ ), et  $\tau$  représente la périodicité de la séquence des matrices d'adjacence des graphes  $G(V(t), E(t), t)$ . Cette formulation est semblable à celle pour un réseau statique simple (voir section 1.5.4), et permet de retrouver le seuil épidémique, qui est atteint lorsque le rayon spectral de  $M$  dépasse la valeur 1. Valdano et al. montrent que le rayon spectral de  $M$  se déduit de celui de la matrice d'accessibilité  $P$  :

$$P = \prod_{i=1}^{T_{max}} 1 - \mu + \lambda A^{(T-i)}$$

Cette matrice rappelle celles évoquées au Chapitre 3 pour l'identification de chemins temporellement cohérents, et est pondérée par des facteurs représentant la probabilité de non-décolonisation ( $1-\mu$ ) et la probabilité d'infection par un voisin ( $\lambda$ ). Son rayon spectral est le même que celui de  $M$ , à une puissance  $1/T_{max}$  près.

### 5.3.2 Perspectives d'application aux données I-Bird

L'adaptation du modèle markovien S-I-S à un substrat dynamique permet de déterminer des seuils épidémiques qui tiennent compte des changements de conformation du réseau des contacts vecteur de la transmission.

#### 5.3.2.1 Evaluation de seuils épidémiques

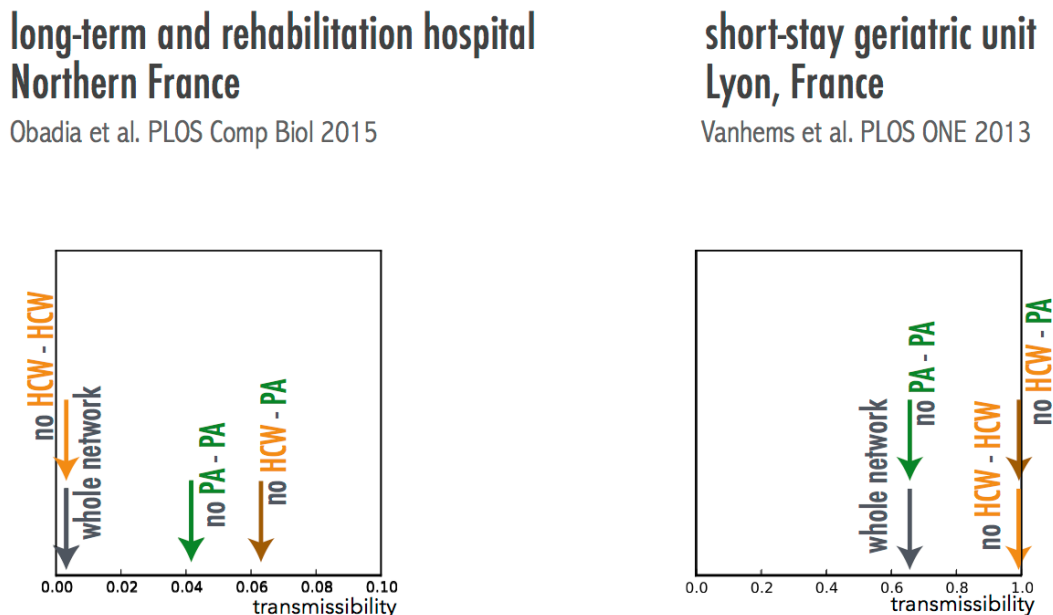
La condition critique  $\rho(M) = 1$ , qui correspond au seuil épidémique, permet le calcul des couples  $(\lambda, \mu)$  qui permettent d'envisager un scénario épidémique compte tenu du réseau des contacts observés. Ces seuils peuvent être étudiés selon plusieurs critères.

##### 5.3.2.1.1 Seuil épidémique selon les contacts

Nous proposons ici des résultats préliminaires, résultats de l'application de la méthode proposée par Valdano et al. Nous comparons le seuil épidémique estimé sur le réseau I-Bird, discrétisé par jour, à celui estimé dans le réseau collecté par Vanhems et al. dans



un service de gériatrie.[144] La Figure 26 montre les valeurs de  $\lambda$ , associées au seuil épidémique critique, pour une valeur  $\mu = 0.1 \text{ jour}^{-1}$  (soit une probabilité de décolonisation journalière de 10%).



**Figure 26. Comparaison du seuil épidémique calculé à partir du réseau des contacts observés dans deux hôpitaux différents.** Les flèches indiquent la valeur du seuil épidémique pour plusieurs graphes. Le réseau entier est indiqué par « whole network ». Les flèches « no HCW-HCW », « no PA-PA » et « no HCW-PA » représentent respectivement les réseaux obtenus en retirant tous les contacts entre personnels de santé, tous les contacts entre patients, et tous les contacts entre un personnel de santé et un patient. Figure en collaboration avec Valdano et al.

Le seuil épidémique évolue différemment selon les contacts qui sont retirés du réseau : dans I-Bird, ce seuil semble peu affecté par les interactions impliquant uniquement les personnels de santé, à l'inverse du réseau de Vanhems et al. Les contacts entre patients et soignants sont par contre source d'un risque accru, comme ceux entre patients (alors que ce n'est pas le cas dans le service de gériatrie).

#### 5.3.2.1.2 Variations dans le temps

La stabilité globale des voisinages, couplée à une loyauté individuelle temps-dépendante, suggère que les seuils épidémiques sont variables dans le temps. Ces données pourront permettre de déterminer des périodes où le risque épidémique est accru, plus stratégiques pour mettre en place des mesures de contrôle.

#### 5.3.2.1.3 Variations dans les différents service de l'hôpital

La charge en soins et la susceptibilité épidémiologique des patients varient selon leur état, et influent sur leur réseau de contacts. Les différents services de l'hôpital peuvent donc présenter des risques épidémiques différents.

### 5.3.2.2 Vers des simulations d'interventions sur la structure du réseau

Les sources de variation des seuils épidémiques sont autant de points permettant de tester des stratégies d'intervention se basant sur le réseau des contacts. Nous raisonnons jusqu'alors en faisant l'hypothèse que le réseau des contacts est également le substrat de la transmission, sur la base des résultats de l'article présenté en section 3.3.2. Nous pouvons dès lors envisager d'altérer le réseau des contacts, afin de réduire le risque épidémique qu'il induit.

Dans la formulation Markovienne de la probabilité d'infection proposée au paragraphe 5.3.1, nous proposons d'introduire un terme de « masque », qui modifie la structure de la matrice  $A^{(t)}$  pour refléter l'impossibilité de certaines transmissions résultant de stratégies de contrôle :

$$P(i, t + 1) = 1 - [1 - (1 - \mu) * P(i, t)] * \prod_{j \neq i} [1 - \lambda(A_{ji}^{(t)} * S_{ji})P(j, t)]$$

$S_{ij}$  représente une terme qui module la possibilité de transmission de  $j$  à  $i$  : le produit  $A_{ji}^{(t)} * S_{ji}$  n'est plus à interpréter comme la réalité d'un contact au sens « interaction sociale », mais comme la résultante épidémiologique de la transmissibilité au cours de cette interaction. La matrice  $S$  est de même dimensions que  $A^{(t)}$ , et permet d'interdire la transmission le long de certaines arêtes de  $A^{(t)}$ , afin de refléter l'impact des mesures de contrôles. Cette matrice doit tenir compte des éventuelles données de colonisation disponibles pour chaque participant, et de différents attributs les concernant (nombre de contacts, durée totale des contacts avec d'autres personnes, catégorie professionnelle, rattachement au même service que  $i$  ou à un autre service...).

L'impact de ces interventions est mesurable par la variation du seuil épidémique, déterminé par la méthode proposée précédemment (paragraphe 5.3.1) : l'augmentation du seuil épidémique indique que les mesures de prévention, proposées par  $S$ , rend la propagation du pathogène plus difficile que dans le réseau original.

## **Conclusion**

Au regard de cette thèse, nous pouvons souligner deux constats majeurs en rapport avec la modélisation épidémiologique. Le premier est que les hypothèses simplificatrices sur les structures des populations ne peuvent pas être supposées vraies dans la plupart des populations « réelles ». Ces hypothèses conduisent souvent à sous-évaluer le risque épidémique,[83, 152] ce qui peut être la source d'une réponse trop tardive, ou moins efficace, en termes de mesures de contrôle pour la santé publique. Le second constat est qu'il est relativement facile de proposer des technologies qui permettent de mesurer de façon très précise l'hétérogénéité des structures de contacts, y compris de façon longitudinale. Les données de l'étude I-Bird sont un bon exemple du « big data » à cet égard.

### **Sur l'hétérogénéité des contacts**

Au cours du Chapitre 1, nous avons réalisé un état de l'art sur l'existence d'un « effet seuil » en épidémiologie, sous plusieurs hypothèses de structures de contacts. Il est possible de trouver des solutions analytiques pour certains modèles épidémiologiques, et ces solutions font intervenir la structure des contacts. En parallèle, nous avons étudié dans le Chapitre 2 le concept des réseaux, permettant de décrire de façon explicite l'hétérogénéité individuelle. Dans le cadre de l'étude I-Bird, nous introduisons une dimension temporelle aux réseaux, en autorisant leur évolution dans le temps. Une approche purement descriptive montre que les contacts interindividuels sont éloignés de l'hypothèse de mélange homogène, avec des propriétés « petit-monde » marquées. Si ces réseaux servent bien de substrat à la dissémination d'un pathogène, les résultats du Chapitre 1 indiquent que cette structure doit être explorée pour refléter la réalité d'une épidémie.

### **Sur la faisabilité de la collecte fine de réseaux de contacts**

Dans le Chapitre 3, nous nous sommes tout d'abord intéressés à l'hypothèse que le réseau des contacts mesurés par des capteurs électroniques de proximité pouvait être utilisés afin d'enregistrer les interactions sociales qui permettent la transmission de *S. aureus* en milieu hospitalier. Nous nous situons donc dans un contexte particulier, où la finalité est de proposer de nouvelles pistes de réflexion pour lutter contre la dissémination de souches multirésistantes, là où les mesures actuelles reposent sur une

sensibilisation collective à l'importance de l'hygiène des mains. Lorsque le portage d'une bactérie multirésistante est avéré chez un patient, des mesures d'isolement peuvent être envisagées. Ces mesures sont contraignantes car elles impliquent un isolement non seulement géographique, mais aussi technique, et une adaptation logistique. Dans le cas où la contamination est réalisée principalement par contacts entre un porteur et un susceptible, des procédures d'intervention visant les contacts, plutôt que le patient lui-même, offrent un nouveau domaine d'étude. Il faut au préalable que ces contacts soient correctement caractérisés.

En développant une méthodologie statistique innovante, nous avons réalisé des simulations de dissémination contrôlées sur le réseau des contacts, et identifié plusieurs quantités mêlant topologie et épidémiologie susceptibles de capturer la corrélation entre les interactions interindividuelles et les événements de transmission de *S. aureus*. L'utilisation de ces mêmes statistiques sur les données de portage collectées dans le cadre du projet I-Bird montrent des tendances similaires, qui reflètent donc une transmission ayant eu lieu le long des arêtes du réseau des contacts. Nous fournissons donc une base théorique permettant de valider l'utilisation d'un proxy électronique pour mesurer les « contacts » au sens épidémiologique, dont nous avons vu que la définition était rarement simple.

### **Sur la caractérisation des contacts « à risque »**

La caractérisation des contacts donnant lieu à transmission est la suite naturelle de cette étude, et occupe le Chapitre 4. Nous utilisons une méthodologie classique en épidémiologie afin de comparer les réseaux des contacts de patients incidents pour SARM à ceux de patients non colonisés. Dans cette analyse cas-témoins, nous identifions des facteurs de risque, qui modulent la probabilité d'acquisition de SARM. Nous avons tout d'abord testé un nombre important de covariables explicatives au cours d'une régression logistique conditionnelle univariée (139 covariables, ayant rapport soit avec des critères démographiques, cliniques, topologiques, ou une combinaison des trois). Les résultats de cette analyse sont cohérents avec la littérature existant sur les modalités de dissémination des bactéries nosocomiales, notamment avec le rôle central du personnel de santé. Les associations observées restaient similaires en intensité lorsque nous ajustons le modèle de régression afin de tenir compte de possibles

facteurs de confusion, comme le fait d'avoir déjà été colonisé par le passé, ou la durée de séjour.

### **Perspectives pour l'élaboration de mesures de contrôle informées par un réseau de contacts**

Au regard des résultats évoqués précédemment, plusieurs pistes de réflexion sont possibles afin de répondre aux problématiques auxquelles nous avons été confrontés lors de cette thèse.

La problématique la plus classique en analyse statistique est celle des données manquantes. Les données de contacts que nous utilisons sont affectées par ce problème, mais nous sommes dans le cas où aucun biais systématique n'est introduit par ce phénomène : le processus d'échantillonnage des contacts (fonctionnement du capteur) est indépendant des processus sociaux qui amènent à la réalisation du réseau. En ce sens, des approches paramétriques comme celles proposées par les ERGM permettent de reconstruire des graphes dont la topologie est similaire à celle observée dans I-Bird. Ces graphes peuvent ensuite être utilisés pour évaluer l'impact de mesures de contrôle.

L'analyse d'un réseau dont la topologie évolue avec le temps est une source de complexité supplémentaire. Nous avons tout d'abord abordé ce concept par une approche « forward », afin de reconstituer des chaînes de transmission temporellement cohérentes. Ce volet de l'analyse était réalisé sur une discrétisation quotidienne. Dans le Chapitre 5, nous proposons, à partir de la méthodologie de Valdano et al., une voie d'étude adaptée à une discrétisation encore plus fine.[142] Des méthodes d'intervention ciblées peuvent être simulées en appliquant un système de « masque » sur les matrices d'adjacence, afin de refléter une modification de la structure des contacts pour un ou quelques individus. L'évaluation de l'impact de différentes mesures pourra passer par la comparaison des seuils épidémiques obtenus sur des réseaux différents, l'objectif étant d'identifier des stratégies qui permettent d'élever ce seuil tout en restant réaliste dans leur possibilité de réalisation.

Nous apportons au fil de cette thèse des éléments qui ouvrent la voie vers des mesures supplémentaires de contrôle, voire de prévention, en matière de dissémination de bactéries multirésistantes. Nous avons utilisé le cas de *S. aureus* résistant à la méticilline,

mais ces approches sont transposables à d'autres bactéries multirésistantes, pour autant que les « contacts » qui les véhiculent soient de nature similaire.

## Bibliographie

1. Albert R, Barabási A-L (2002) Statistical mechanics of complex networks. *Rev Mod Phys* 74:47–97. doi: 10.1103/RevModPhys.74.47
2. Albert R, Jeong H, Barabási A-L (1999) Internet: Diameter of the World-Wide Web. *Nature* 401:130–131. doi: 10.1038/43601
3. Albrich WC, Harbarth S (2008) Health-care workers: source, vector, or victim of MRSA? *Lancet Infect Dis* 8:289–301. doi: 10.1016/S1473-3099(08)70097-5
4. Aldous JM, Wilson RJ (2000) *Graphs and Applications: An Introductory Approach*. Springer Science & Business Media
5. Anderson RM, May RM (1992) *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press
6. Balcan D, Colizza V, Gonçalves B, Hu H, Ramasco JJ, Vespignani A (2009) Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc Natl Acad Sci* 106:21484–21489. doi: 10.1073/pnas.0906910106
7. Bansal S, Grenfell BT, Meyers LA (2007) When individual behaviour matters: homogeneous and network models in epidemiology. *J R Soc Interface* 4:879–891. doi: 10.1098/rsif.2007.1100
8. Barabási A-L, Albert R (1999) Emergence of Scaling in Random Networks. *Science* 286:509–512. doi: 10.1126/science.286.5439.509
9. Barber M (1961) Methicillin-resistant staphylococci. *J Clin Pathol* 14:385–393. doi: 10.1136/jcp.14.4.385
10. Barclay VC, Smieszek T, He J, Cao G, Rainey JJ, Gao H, Uzicanin A, Salathé M (2014) Positive Network Assortativity of Influenza Vaccination at a High School: Implications for Outbreak Risk and Herd Immunity. *PLoS ONE* 9:e87042. doi: 10.1371/journal.pone.0087042
11. Barrat A, Barthélemy M, Vespignani A (2008) *Dynamical Processes on Complex Networks*. Cambridge University Press
12. Bellman R (1958) On a Routing Problem. *Q Appl Math* 16:87–90.
13. Bettencourt LMA, Ribeiro RM (2008) Real time bayesian estimation of the epidemic potential of emerging infectious diseases. *PLoS One* 3:e2185.
14. Blok HEM, Troelstra A, Kamp-Hopmans TEM, Gigengack-Baars ACM,

- Vandenbroucke-Grauls CMJE, Weersink AJL, Verhoef J, Mascini EM (2003) Role of Healthcare Workers in Outbreaks of Methicillin-Resistant *Staphylococcus aureus*: A 10-year Evaluation From a Dutch University Hospital. *Infect Control Hosp Epidemiol* 24:679–685. doi: 10.1086/502275
15. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech Theory Exp* 2008:P10008. doi: 10.1088/1742-5468/2008/10/P10008
  16. Boguñá M, Castellano C, Pastor-Satorras R (2013) Nature of the Epidemic Threshold for the Susceptible-Infected-Susceptible Dynamics in Networks. *Phys Rev Lett* 111:068701. doi: 10.1103/PhysRevLett.111.068701
  17. Boucher HW, Corey GR (2008) Epidemiology of Methicillin-Resistant *Staphylococcus aureus*. *Clin Infect Dis* 46:S344–S349. doi: 10.1086/533590
  18. Burnet SFM (1940) *Biological Aspects of Infectious Disease*. The University Press
  19. Castellano C, Pastor-Satorras R (2010) Thresholds for Epidemic Spreading in Networks. *Phys Rev Lett* 105:218701. doi: 10.1103/PhysRevLett.105.218701
  20. Cattuto C, Van den Broeck W, Barrat A, Colizza V, Pinton J-F, Vespignani A (2010) Dynamics of Person-to-Person Interactions from Distributed RFID Sensor Networks. *PLoS ONE* 5:e11596. doi: 10.1371/journal.pone.0011596
  21. Cauchemez S, Bhattarai A, Marchbanks TL, Fagan RP, Ostroff S, Ferguson NM, Swerdlow D, Sodha SV, Moll ME, Angulo FJ, Palekar R, Archer WR, Finelli L (2011) Role of Social Networks in Shaping Disease Transmission During a Community Outbreak of 2009 H1N1 Pandemic Influenza. *Proc Natl Acad Sci* 108:2825–2830. doi: 10.1073/pnas.1008895108
  22. Cauchemez S, Boëlle P-Y, Donnelly CA, Ferguson NM, Thomas G, Leung GM, Hedley AJ, Anderson RM, Valleron A-J (2006) Real-time Estimates in Early Detection of SARS. *Emerg Infect Dis* 12:110–113. doi: 10.3201/eid1201.050593
  23. Christley RM, Pinchbeck GL, Bowers RG, Clancy D, French NP, Bennett R, Turner J (2005) Infection in Social Networks: Using Network Analysis to Identify High-Risk Individuals. *Am J Epidemiol* 162:1024–1031. doi: 10.1093/aje/kwi308
  24. Chung F, Lu L, Vu V (2003) Spectra of random graphs with given expected degrees. *Proc Natl Acad Sci* 100:6313–6318. doi: 10.1073/pnas.0937490100
  25. Cohen ML (1992) Epidemiology of Drug Resistance: Implications for a Post—Antimicrobial Era. *Science* 257:1050–1055. doi: 10.1126/science.257.5073.1050



26. Cohen R, Havlin S (2003) Scale-Free Networks Are Ultrasmall. *Phys Rev Lett* 90:058701. doi: 10.1103/PhysRevLett.90.058701
27. Colizza V, Vespignani A (2007) Invasion Threshold in Heterogeneous Metapopulation Networks. *Phys Rev Lett* 99:148701. doi: 10.1103/PhysRevLett.99.148701
28. Cookson B, Peters B, Webster M, Phillips I, Rahman M, Noble W (1989) Staff carriage of epidemic methicillin-resistant *Staphylococcus aureus*. *J Clin Microbiol* 27:1471–1476.
29. Cori A, Valleron AJ, Carrat F, Scalia Tomba G, Thomas G, Boëlle PY (2012) Estimating influenza latency and infectious period durations using viral excretion data. *Epidemics* 4:132–138. doi: 10.1016/j.epidem.2012.06.001
30. Cosgrove SE, Sakoulas G, Perencevich EN, Schwaber MJ, Karchmer AW, Carmeli Y (2003) Comparison of Mortality Associated with Methicillin-Resistant and Methicillin-Susceptible *Staphylococcus aureus* Bacteremia: A Meta-analysis. *Clin Infect Dis* 36:53–59. doi: 10.1086/345476
31. Couderc C, Jolivet S, Thiébaut ACM, Ligier C, Remy L, Alvarez A-S, Lawrence C, Salomon J, Herrmann J-L, Guillemot D, Bernède-Bauduin C, Bertucci I, Dupont C, Minor OL, Petit A, Sorel T (2014) Fluoroquinolone Use Is a Risk Factor for Methicillin-Resistant *Staphylococcus aureus* Acquisition in Long-term Care Facilities: A Nested Case-Case-Control Study. *Clin Infect Dis* 59:206–215. doi: 10.1093/cid/ciu236
32. Cutler RR (1979) Relationship Between Antibiotic Resistance, the Production of “Virulence Factors”, and Virulence for Experimental Animals in *Staphylococcus Aureus*. *J Med Microbiol* 12:55–62. doi: 10.1099/00222615-12-1-55
33. Dall’Antonia M, Coen PG, Wilks M, Whiley A, Millar M (2005) Competition between methicillin-sensitive and -resistant *Staphylococcus aureus* in the anterior nares. *J Hosp Infect* 61:62–67. doi: 10.1016/j.jhin.2005.01.008
34. Danon L, Ford AP, House T, Jewell CP, Keeling MJ, Roberts GO, Ross JV, Vernon MC (2011) Networks and the Epidemiology of Infectious Disease. *Interdiscip Perspect Infect Dis* 2011:e284909. doi: 10.1155/2011/284909
35. Ben-David D, Mermel LA, Parenteau S (2008) Methicillin-resistant *Staphylococcus aureus* transmission: The possible importance of unrecognized health care worker carriage. *Am J Infect Control* 36:93–97. doi: 10.1016/j.ajic.2007.05.013
36. Diekmann O, Heesterbeek JAP (2000) *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. John Wiley & Sons

37. Dijkstra EW (1959) A note on two problems in connexion with graphs. *Numer Math* 1:269–271. doi: 10.1007/BF01386390
38. Doherty IA, Padian NS, Marlow C, Aral SO (2005) Determinants and Consequences of Sexual Networks as They Affect the Spread of Sexually Transmitted Infections. *J Infect Dis* 191:S42–S54. doi: 10.1086/425277
39. Donker T, Wallinga J, Slack R, Grundmann H (2012) Hospital Networks and the Dispersal of Hospital-Acquired Pathogens by Patient Transfer. *PLoS ONE*. doi: 10.1371/journal.pone.0035002
40. Van Druten J, Reintjes A, Plantinga A (1990) Mumps, measles and rubella, a longitudinal serological study into degree of protection and risk of infection. (In Dutch). Hague Neth. Praeventiefonds
41. Dubé C, Ribble C, Kelton D, McNab B (2009) A Review of Network Analysis Terminology and its Application to Foot-and-Mouth Disease Modelling and Policy Development. *Transbound Emerg Dis* 56:73–85. doi: 10.1111/j.1865-1682.2008.01064.x
42. Eames K, Bansal S, Frost S, Riley S (2015) Six challenges in measuring contact networks for use in modelling. *Epidemics* 10:72–77. doi: 10.1016/j.epidem.2014.08.006
43. Eames KTD, Keeling MJ (2004) Monogamous networks and the spread of sexually transmitted diseases. *Math Biosci* 189:115–130. doi: 10.1016/j.mbs.2004.02.003
44. Eckmanns T, Schwab F, Bessert J, Wettstein R, Behnke M, Grundmann H, Rüden H, Gastmeier P (2006) Hand rub consumption and hand hygiene compliance are not indicators of pathogen transmission in intensive care units. *J Hosp Infect* 63:406–411. doi: 10.1016/j.jhin.2006.03.015
45. Edmunds WJ, O'callaghan CJ, Nokes DJ (1997) Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections. *Proc R Soc Lond B Biol Sci* 264:949–957. doi: 10.1098/rspb.1997.0131
46. Eggo RM, Cauchemez S, Ferguson NM (2011) Spatial dynamics of the 1918 influenza pandemic in England, Wales and the United States. *J R Soc Interface* 8:233–243. doi: 10.1098/rsif.2010.0216
47. Erdős P, Rényi A (1959) On random graphs. *Publ Math Debr* 6:290–297.
48. Eubank S, Guclu H, Anil Kumar VS, Marathe MV, Srinivasan A, Toroczkai Z, Wang N (2004) Modelling disease outbreaks in realistic urban social networks. *Nature* 429:180–184. doi: 10.1038/nature02541

49. European Center for Disease Prevention and Control, European Medicines Agency (2009) The bacterial challenge: time to react.
50. Eveillard M, Hitoto H, Raymond F, Kouatchet A, Dubé L, Guilloteau V, Pradelle M-T, Brunel P, Mercat A, Joly-Guillou M-L (2009) Measurement and interpretation of hand hygiene compliance rates: importance of monitoring entire care episodes. *J Hosp Infect* 72:211–217. doi: 10.1016/j.jhin.2009.03.025
51. Farrington CP, Kanaan MN, Gay NJ (2001) Estimation of the basic reproduction number for infectious diseases from age-stratified serological survey data. *J R Stat Soc Ser C Appl Stat* 50:251–292. doi: 10.1111/1467-9876.00233
52. Ferguson NM, Cummings DAT, Cauchemez S, Fraser C, Riley S, Meeyai A, Iamsirithaworn S, Burke DS (2005) Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* 437:209–214. doi: 10.1038/nature04017
53. Fleming A (1929) On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of *B. influenzae*. *Br J Exp Pathol* 10:226–236.
54. Ford LR (1956) Network flow theory. Defense Technical Information Center
55. Forsberg White L, Pagano M (2008) A likelihood-based method for real-time estimation of the serial interval and reproductive number of an epidemic. *Stat Med* 27:2999–3016. doi: 10.1002/sim.3136
56. Frank O, Strauss D (1986) Markov Graphs. *J Am Stat Assoc* 81:832–842. doi: 10.1080/01621459.1986.10478342
57. Frénay HME, Bunschoten AE, Schouls LM, Leeuwen WJ van, Vandenbroucke-Grauls CMJE, Verhoef J, Mooi FR (1996) Molecular typing of methicillin-resistant *Staphylococcus aureus* on the basis of protein A gene polymorphism. *Eur J Clin Microbiol Infect Dis* 15:60–64. doi: 10.1007/BF01586186
58. Gardy JL, Johnston JC, Sui SJH, Cook VJ, Shah L, Brodtkin E, Rempel S, Moore R, Zhao Y, Holt R, Varhol R, Birol I, Lem M, Sharma MK, Elwood K, Jones SJM, Brinkman FSL, Brunham RC, Tang P (2011) Whole-Genome Sequencing and Social-Network Analysis of a Tuberculosis Outbreak. *N Engl J Med* 364:730–739. doi: 10.1056/NEJMoa1003176
59. Ghani AC, Swinton J, Garnett GP (1997) The role of sexual partnership networks in the epidemiology of gonorrhea. *Sex Transm Dis* 24:45–56.
60. Gómez S, Arenas A, Borge-Holthoefer J, Meloni S, Moreno Y (2010) Discrete-time Markov chain approach to contact-based disease spreading in complex networks. *EPL*

Europhys Lett 89:38009. doi: 10.1209/0295-5075/89/38009

61. González MC, Hidalgo CA, Barabási A-L (2008) Understanding individual human mobility patterns. *Nature* 453:779–782. doi: 10.1038/nature06958
62. Goodreau SM (2007) Advances in exponential random graph ( $p^*$ ) models applied to a large social network. *Soc Netw* 29:231–248. doi: 10.1016/j.socnet.2006.08.001
63. Goodreau SM, Kitts JA, Morris M (2009) Birds of a feather, or friend of a friend? using exponential random graph models to investigate adolescent social networks\*. *Demography* 46:103–125. doi: 10.1353/dem.0.0045
64. Grais RF, Ellis JH, Glass GE (2003) Assessing the impact of airline travel on the geographic spread of pandemic influenza. *Eur J Epidemiol* 18:1065–1072. doi: 10.1023/A:1026140019146
65. Grassly NC, Fraser C (2008) Mathematical models of infectious disease transmission. *Nat Rev Microbiol* 6:477–487. doi: 10.1038/nrmicro1845
66. Gumel AB, Ruan S, Day T, Watmough J, Brauer F, Driessche P van den, Gabrielson D, Bowman C, Alexander ME, Ardal S, Wu J, Sahai BM (2004) Modelling strategies for controlling SARS outbreaks. *Proc R Soc Lond B Biol Sci* 271:2223–2232. doi: 10.1098/rspb.2004.2800
67. Halloran ME, Longini IM, Nizam A, Yang Y (2002) Containing Bioterrorist Smallpox. *Science* 298:1428–1432. doi: 10.1126/science.1074674
68. Handcock MS, Gile KJ (2010) Modeling social networks from sampled data. *Ann Appl Stat* 4:5–25. doi: 10.1214/08-AOAS221
69. Handcock MS, Hunter DR, Butts CT, Goodreau SM, Morris M (2008) statnet: Software tools for the representation, visualization, analysis and simulation of network data. *J Stat Softw* 24:1548.
70. Harris AD, Samore MH, Lipsitch M, Kaye KS, Perencevich E, Carmeli Y (2002) Control-Group Selection Importance in Studies of Antimicrobial Resistance: Examples Applied to *Pseudomonas aeruginosa*, *Enterococci*, and *Escherichia coli*. *Clin Infect Dis* 34:1558–1563. doi: 10.1086/340533
71. Hartman BJ, Tomasz A (1986) Expression of methicillin resistance in heterogeneous strains of *Staphylococcus aureus*. *Antimicrob Agents Chemother* 29:85–92. doi: 10.1128/AAC.29.1.85
72. Haydon DT, Chase-Topping M, Shaw DJ, Matthews L, Friar JK, Wilesmith J, Woolhouse MEJ (2003) The construction and analysis of epidemic trees with reference

- to the 2001 UK foot-and-mouth outbreak. *Proc R Soc Lond B Biol Sci* 270:121–127. doi: 10.1098/rspb.2002.2191
73. Hobgood C, Villani J, Quattlebaum R (2005) Impact of Emergency Department Volume on Registered Nurse Time at the Bedside. *Ann Emerg Med* 46:481–489. doi: 10.1016/j.annemergmed.2005.07.014
74. Hollingsworth JC, Chisholm CD, Giles BK, Cordell WH, Nelson DR (1998) How Do Physicians and Nurses Spend Their Time in the Emergency Department? *Ann Emerg Med* 31:87–91. doi: 10.1016/S0196-0644(98)70287-2
75. Holme P, Saramäki J (2012) Temporal networks. *Phys Rep* 519:97–125. doi: 10.1016/j.physrep.2012.03.001
76. Hornbeck T, Naylor D, Segre AM, Thomas G, Herman T, Polgreen PM (2012) Using Sensor Networks to Study the Effect of Peripatetic Healthcare Workers on the Spread of Hospital-Associated Infections. *J Infect Dis* 206:1549–1557. doi: 10.1093/infdis/jis542
77. Hufnagel L, Brockmann D, Geisel T (2004) Forecast and control of epidemics in a globalized world. *Proc Natl Acad Sci U S A* 101:15124–15129. doi: 10.1073/pnas.0308344101
78. Hunter DR, Handcock MS (2006) Inference in Curved Exponential Family Models for Networks. *J Comput Graph Stat* 15:565–583. doi: 10.1198/106186006X133069
79. Isella L, Romano M, Barrat A, Cattuto C, Colizza V, Van den Broeck W, Gesualdo F, Pandolfi E, Ravà L, Rizzo C, Tozzi AE (2011) Close Encounters in a Pediatric Ward: Measuring Face-to-Face Proximity and Mixing Patterns with Wearable Sensors. *PLoS ONE* 6:e17144. doi: 10.1371/journal.pone.0017144
80. Jevons MP, Coe AW, Parker MT (1963) Methicillin resistance in staphylococci. *Lancet* 1:904–907.
81. Kardaś-Słoma L, Boëlle PY, Opatowski L, Brun-Buisson C, Guillemot D, Temime L (2011) Impact of Antibiotic Exposure Patterns on Selection of Community-Associated Methicillin-Resistant *Staphylococcus aureus* in Hospital Settings. *Antimicrob Agents Chemother* 55:4888–4895. doi: 10.1128/AAC.01626-10
82. Keeling MJ (1999) The effects of local spatial structure on epidemiological invasions. *Proc R Soc Lond B Biol Sci* 266:859–867. doi: 10.1098/rspb.1999.0716
83. Keeling MJ, Eames KTD (2005) Networks and epidemic models. *J R Soc Interface* 2:295–307. doi: 10.1098/rsif.2005.0051
84. Keeling MJ, Rohani P (2008) *Modeling Infectious Diseases in Humans and Animals*.

Princeton University Press

85. Kermack WO, McKendrick AG (1927) A Contribution to the Mathematical Theory of Epidemics. *Proc R Soc Lond Math Phys Eng Sci* 115:700–721. doi: 10.1098/rspa.1927.0118
86. Kermack WO, McKendrick AG (1932) Contributions to the Mathematical Theory of Epidemics. II. The Problem of Endemicity. *Proc R Soc Lond Math Phys Eng Sci* 138:55–83. doi: 10.1098/rspa.1932.0171
87. Kermack WO, McKendrick AG (1933) Contributions to the Mathematical Theory of Epidemics. III. Further Studies of the Problem of Endemicity. *Proc R Soc Lond Math Phys Eng Sci* 141:94–122. doi: 10.1098/rspa.1933.0106
88. Kinsman OS, Naidoo J, Noble WC (1985) Some effects of plasmids coding for antibiotic resistance on the virulence of *Staphylococcus aureus*. *Br J Exp Pathol* 66:325–332.
89. Kirby WMM (1944) Extraction of a Highly Potent Penicillin Inactivator from Penicillin Resistant *Staphylococci*. *Science* 99:452–453. doi: 10.1126/science.99.2579.452
90. Klov Dahl AS (1985) Social networks and the spread of infectious diseases: The AIDS example. *Soc Sci Med* 21:1203–1216. doi: 10.1016/0277-9536(85)90269-2
91. Kretzschmar M, Duynhoven YTHP van, Severijnen AJ (1996) Modeling Prevention Strategies for Gonorrhea and Chlamydia Using Stochastic Network Simulations. *Am J Epidemiol* 144:306–317.
92. Krivitsky PN, Handcock MS (2014) A separable model for dynamic networks. *J R Stat Soc Ser B Stat Methodol* 76:29–46. doi: 10.1111/rssb.12014
93. Kuehnert MJ, Hill HA, Kupronis BA, Tokars JI, Solomon SL, Jernigan DB (2005) Methicillin-resistant-*Staphylococcus aureus* Hospitalizations, United States. *Emerg Infect Dis* 11:468–872. doi: 10.3201/eid1106.040831
94. Laupland KB, Church DL, Mucenski M, Sutherland LR, Davies HD (2003) Population-Based Study of the Epidemiology of and the Risk Factors for Invasive *Staphylococcus aureus* Infections. *J Infect Dis* 187:1452–1459. doi: 10.1086/374621
95. Lau T, Tang G, Mak K, Leung G (2014) Moment-specific compliance with hand hygiene. *Clin Teach* 11:159–164. doi: 10.1111/tct.12088
96. Lentz HHK, Selhorst T, Sokolov IM (2013) Unfolding Accessibility Provides a Macroscopic Approach to Temporal Networks. *Phys Rev Lett* 110:118701. doi:

10.1103/PhysRevLett.110.118701

97. Lessler J, Reich NG, Cummings DAT (2009) Outbreak of 2009 pandemic influenza A (H1N1) at a New York City school. *N Engl J Med* 361:2628–2636.
98. Lowy FD (2003) Antimicrobial resistance: the example of *Staphylococcus aureus*. *J Clin Invest* 111:1265–1273. doi: 10.1172/JCI200318535
99. Lucet J-C, Laouenan C, Chelius G, Veziris N, Lepelletier D, Friggeri A, Abiteboul D, Bouvet E, Mentré F, Fleury E (2012) Electronic sensors for assessing interactions between healthcare workers and patients under airborne precautions. *PloS One* 7:e37893. doi: 10.1371/journal.pone.0037893
100. Lucet J-C, Paoletti X, Demontpion C, Degrave M, Vanjak D, Vincent C, Andreumont A, Jarlier V, Mentré F, Nicolas-Chanoine M-H (2009) Carriage of methicillin-resistant *staphylococcus aureus* in home care settings: Prevalence, duration, and transmission to household members. *Arch Intern Med* 169:1372–1378. doi: 10.1001/archinternmed.2009.217
101. Meyers LA, Pourbohloul B, Newman MEJ, Skowronski DM, Brunham RC (2005) Network theory and SARS: predicting outbreak diversity. *J Theor Biol* 232:71–81. doi: 10.1016/j.jtbi.2004.07.026
102. Miller JC (2009) Spread of infectious disease through clustered populations. *J R Soc Interface* rsif.2008.0524. doi: 10.1098/rsif.2008.0524
103. Moody J (2002) The Importance of Relationship Timing for Diffusion. *Soc Forces* 81:25–56. doi: 10.1353/sof.2002.0056
104. Moorman DR, Mandell GL (1981) Characteristics of rifampin-resistant variants obtained from clinical isolates of *Staphylococcus aureus*. *Antimicrob Agents Chemother* 20:709–713.
105. Morris M (1997) Sexual networks and HIV. *AIDS Lond Engl* 11 Suppl A:S209–216.
106. Morris M (2004) *Network Epidemiology: A Handbook for Survey Design and Data Collection: A Handbook for Survey Design and Data Collection*. OUP Oxford
107. Morris M, Handcock MS, Hunter DR (2008) Specification of Exponential-Family Random Graph Models: Terms and Computational Aspects. *J Stat Softw* 24:1548–7660.
108. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, Massari M, Salmaso S, Tomba GS, Wallinga J, Heijne J, Sadkowska-Todys M, Rosinska M, Edmunds WJ (2008) Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLoS Med* 5:e74. doi: 10.1371/journal.pmed.0050074

109. Mylius SD, Hagenaars TJ, Lugnér AK, Wallinga J (2008) Optimal allocation of pandemic influenza vaccine depends on age, risk and timing. *Vaccine* 26:3742–3749. doi: 10.1016/j.vaccine.2008.04.043
110. Mylotte JM, McDermott C, Spooner JA (1987) Prospective study of 114 consecutive episodes of *Staphylococcus aureus* bacteremia. *Rev Infect Dis* 9:891–907.
111. Newman M (2010) *Networks: An Introduction*, 1 edition. Oxford University Press, Oxford; New York
112. Newman MEJ (2002) Spread of epidemic disease on networks. *Phys Rev E* 66:016128. doi: 10.1103/PhysRevE.66.016128
113. Noble JV (1974) Geographic and temporal development of plagues. *Nature* 250:726–729. doi: 10.1038/250726a0
114. Obadia T, Haneef R, Boëlle P-Y (2012) The R0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks. *BMC Med Inform Decis Mak* 12:147. doi: 10.1186/1472-6947-12-147
115. Pastor-Satorras R, Vespignani A (2001) Epidemic Spreading in Scale-Free Networks. *Phys Rev Lett* 86:3200–3203. doi: 10.1103/PhysRevLett.86.3200
116. Pattison P, Wasserman S (1999) Logit models and logistic regressions for social networks: II. Multivariate relations. *Br J Math Stat Psychol* 52:169–193. doi: 10.1348/000711099159053
117. Pittet D, Allegranzi B, Sax H, Dharan S, Pessoa-Silva CL, Donaldson L, Boyce JM (2006) Evidence-based model for hand transmission during patient care and the role of improved practices. *Lancet Infect Dis* 6:641–652. doi: 10.1016/S1473-3099(06)70600-4
118. Pittet D, Allegranzi B, Storr J, Donaldson L (2006) “Clean Care is Safer Care”: the Global Patient Safety Challenge 2005–2006. *Int J Infect Dis* 10:419–424. doi: 10.1016/j.ijid.2006.06.001
119. Pittet D, MD, MS, Harbarth S MD, Ruef C MD, Francioli P MD, Sudre P, MD, MS, Pétignat C MD, Trampuz A MD, Widmer A, MD, MS (1999) Prevalence and Risk Factors for Nosocomial Infections in Four University Hospitals in Switzerland •. *Infect Control Hosp Epidemiol* 20:37–42. doi: 10.1086/501554
120. Pittet D, Simon A, Hugonnet S, Pessoa-Silva CL, Sauvan V, Perneger TV (2004) Hand Hygiene among Physicians: Performance, Beliefs, and Perceptions. *Ann Intern Med* 141:1–8. doi: 10.7326/0003-4819-141-1-200407060-00008
121. Riley S, Fraser C, Donnelly CA, Ghani AC, Abu-Raddad LJ, Hedley AJ, Leung GM, Ho L-



- M, Lam T-H, Thach TQ, Chau P, Chan K-P, Lo S-V, Leung P-Y, Tsang T, Ho W, Lee K-H, Lau EMC, Ferguson NM, Anderson RM (2003) Transmission Dynamics of the Etiological Agent of SARS in Hong Kong: Impact of Public Health Interventions. *Science* 300:1961–1966. doi: 10.1126/science.1086478
122. Robins G, Pattison P, Kalish Y, Lusher D (2007) An introduction to exponential random graph (p\*) models for social networks. *Soc Netw* 29:173–191. doi: 10.1016/j.socnet.2006.08.002
123. Robins G, Pattison P, Wasserman S (1999) Logit models and logistic regressions for social networks: III. Valued relations. *Psychometrika* 64:371–394. doi: 10.1007/BF02294302
124. Robinson K, Cohen T, Colijn C (2012) The dynamics of sexual contact networks: Effects on disease spread and control. *Theor Popul Biol* 81:89–96. doi: 10.1016/j.tpb.2011.12.009
125. Rocha LEC, Liljeros F, Holme P (2010) Information dynamics shape the sexual networks of Internet-mediated prostitution. *Proc Natl Acad Sci* 107:5706–5711. doi: 10.1073/pnas.0914080107
126. Roethlisberger FJ, Dickson WJ (1949) *Management and the Worker*. (1939) 9th Pr. Harvard University Press
127. Ross R (1916) An Application of the Theory of Probabilities to the Study of a priori Pathometry. Part I. *Proc R Soc Lond Math Phys Eng Sci* 92:204–230. doi: 10.1098/rspa.1916.0007
128. Rvachev LA, Longini IMJ (1985) A mathematical model for the global spread of influenza. *Math Biosci* 75:3–22. doi: 10.1016/0025-5564(85)90064-1
129. Salathé M, Bengtsson L, Bodnar TJ, Brewer DD, Brownstein JS, Buckee C, Campbell EM, Cattuto C, Khandelwal S, Mabry PL, Vespignani A (2012) Digital Epidemiology. *PLoS Comput Biol* 8:e1002616. doi: 10.1371/journal.pcbi.1002616
130. Salathé M, Kazandjieva M, Lee JW, Levis P, Feldman MW, Jones JH (2010) A High-Resolution Human Contact Network for Infectious Disease Transmission. *Proc Natl Acad Sci* 107:22020–22025. doi: 10.1073/pnas.1009094108
131. Shenoy ES, Paras ML, Noubary F, Walensky RP, Hooper DC (2014) Natural history of colonization with methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant *Enterococcus* (VRE): a systematic review. *BMC Infect Dis* 14:177. doi: 10.1186/1471-2334-14-177

132. Smieszek T, Barclay VC, Seeni I, Rainey JJ, Gao H, Uzicanin A, Salathé M (2014) How should social mixing be measured: comparing web-based survey and sensor-based methods. *BMC Infect Dis* 14:136. doi: 10.1186/1471-2334-14-136
133. Smieszek T, Salathé M (2013) A low-cost method to assess the epidemiological importance of individuals in controlling infectious disease outbreaks. *BMC Med* 11:35. doi: 10.1186/1741-7015-11-35
134. Snijders TAB (2002) Markov Chain Monte Carlo Estimation of Exponential Random Graph Models. *J. Soc. Struct.* 3:
135. Swartz MN (1997) Use of Antimicrobial Agents and Drug Resistance. *N Engl J Med* 337:491–492. doi: 10.1056/NEJM199708143370709
136. Tacconelli E, Angelis GD, Cataldo MA, Pozzi E, Cauda R (2008) Does antibiotic exposure increase the risk of methicillin-resistant *Staphylococcus aureus* (MRSA) isolation? A systematic review and meta-analysis. *J Antimicrob Chemother* 61:26–38. doi: 10.1093/jac/dkm416
137. Temime L, Opatowski L, Pannet Y, Brun-Buisson C, Boëlle PY, Guillemot D (2009) Peripatetic Health-Care Workers as Potential Superspreaders. *Proc Natl Acad Sci* 106:18420–18425. doi: 10.1073/pnas.0900974106
138. Thompson SK, Frank O (2000) Model-based estimation with link-tracing sampling designs. *Surv Methodol* 26:87–98.
139. Thompson SK, Seber GAF (1996) Adaptive sampling. Wiley
140. Ueno T, Masuda N (2008) Controlling nosocomial infection based on structure of hospital social networks. *J Theor Biol* 254:655–666. doi: 10.1016/j.jtbi.2008.07.001
141. Utsui Y, Yokota T (1985) Role of an altered penicillin-binding protein in methicillin- and cephem-resistant *Staphylococcus aureus*. *Antimicrob Agents Chemother* 28:397–403. doi: 10.1128/AAC.28.3.397
142. Valdano E, Ferreri L, Poletto C, Colizza V (2015) Analytical Computation of the Epidemic Threshold on Temporal Networks. *Phys Rev X* 5:021005. doi: 10.1103/PhysRevX.5.021005
143. Valdano E, Poletto C, Giovannini A, Palma D, Savini L, Colizza V (2015) Predicting Epidemic Risk from Past Temporal Contact Data. *PLoS Comput Biol* 11:e1004152. doi: 10.1371/journal.pcbi.1004152
144. Vanhems P, Barrat A, Cattuto C, Pinton J-F, Khanafer N, Régis C, Kim B, Comte B, Voirin N (2013) Estimating Potential Infection Transmission Routes in Hospital Wards

- Using Wearable Proximity Sensors. PLoS ONE 8:e73970. doi: 10.1371/journal.pone.0073970
145. Wallinga J, Lévy-Bruhl D, Gay NJ, Wachmann CH (2001) Estimation of measles reproduction ratios and prospects for elimination of measles by vaccination in some Western European countries. *Epidemiol Infect* 127:281–295.
  146. Wallinga J, Lipsitch M (2007) How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc R Soc B Biol Sci* 274:599.
  147. Wallinga J, Teunis P (2004) Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am J Epidemiol* 160:509.
  148. Wallinga J, Teunis P, Kretzschmar M (2006) Using Data on Social Contacts to Estimate Age-specific Transmission Parameters for Respiratory-spread Infectious Agents. *Am J Epidemiol* 164:936–944. doi: 10.1093/aje/kwj317
  149. Walsh TJ, Vlahov D, Hansen SL, Sonnenberg E, Khabbaz R, Gadacz T, Standiford HC (1987) Prospective microbiologic surveillance in control of nosocomial methicillin-resistant *Staphylococcus aureus*. *Infect Control IC* 8:7–14.
  150. Wang Y, Chakrabarti D, Wang C, Faloutsos C (2003) Epidemic spreading in real networks: an eigenvalue viewpoint. 22nd Int. Symp. Reliab. Distrib. Syst. 2003 Proc. pp 25–34
  151. Wasserman S, Pattison P (1996) Logit models and logistic regressions for social networks: I. An introduction to Markov graphs andp. *Psychometrika* 61:401–425. doi: 10.1007/BF02294547
  152. Watts DJ, Strogatz SH (1998) Collective dynamics of “small-world” networks. *Nature* 393:440–442. doi: 10.1038/30918
  153. Whitby M, McLaws ML, Berry G (2001) Risk of death from methicillin-resistant *Staphylococcus aureus* bacteraemia: a meta-analysis. *Med J Aust* 175:264–267.
  154. Woodhouse DE, Rothenberg RB, Potterat JJ, Darrow WW, Muth SQ, Klovdahl AS, Zimmerman HP, Rogers HL, Maldonado TS, Muth JB (1994) Mapping a social network of heterosexuals at high risk for HIV infection. *AIDS Lond Engl* 8:1331–1336.
  155. World Health Organization (2005) World Alliance for Patient Safety: Global Patient Safety Challenge 2005-2006.
  156. Xia Y, Bjørnstad ON, Grenfell BT (2004) Measles Metapopulation Dynamics: A Gravity Model for Epidemiological Coupling and Dynamics. *Am Nat* 164:267–281. doi: 10.1086/422341

157. Zierdt CH (1982) Long-term *Staphylococcus aureus* carrier state in hospital patients. J Clin Microbiol 16:517–520.

# Annexe 1 : Résultats de l'analyse cas-témoins pour la colonisation des patients par SARM en utilisant des données lissées

Un dictionnaire des covariables est proposé en Annexe 3.

Univariate - SA_type = SARM - Level = 1 - Lag = 7 - Duration = 300						
	N	CASE (N=93 ) min/mean/max	CTRL (N=669 ) min/mean/max	Test Statistic	SD	Odds Ratio per unit
age	762	25/59/86	27/59.1/103	Z=0.18 d.f.=1 P=0.861	16.3	1.001 [ 0.987 ; 1.015 ]
sexe	762	LogRank Test	Overall effect	F=0.75 d.f.=1 P=0.3862		NA
sexe.M	762	50.5	48	Z=0.87 d.f.=NA, P=0.3867	NA	1.213 [ 0.783 ; 1.878 ]
mif	762	LogRank Test	Overall effect	F=5.74 d.f.=3 P=0.1247		NA
mif.[21,42]	762	29	25.3	Z=2.08 d.f.=NA, P=0.0379	NA	2.067 [ 1.042 ; 4.1 ]
mif.(42,85]	762	30.1	25.7	Z=2 d.f.=NA, P=0.0457	NA	2.016 [ 1.014 ; 4.009 ]
mif.(85,118]	762	22.6	23.3	Z=1.12 d.f.=NA, P=0.2615	NA	1.516 [ 0.733 ; 3.133 ]
time.at.risk	761	5/62.9/144	4/54.9/148	Z=2.07 d.f.=1 P=0.0384	42.6	1.006 [ 1 ; 1.011 ]
atb.expo	762	LogRank Test	Overall effect	F=8.4 d.f.=1 P=0.0038		NA
atb.expo.TRUE	762	15.1	6.6	Z=3.12 d.f.=NA, P=0.0018	NA	2.776 [ 1.462 ; 5.268 ]
atb.mssa.sensi	762	LogRank Test	Overall effect	F=13.42 d.f.=2 P=0.0012		NA
atb.mssa.sensi.R	762	4.3	0.9	Z=2.65 d.f.=NA, P=0.0081	NA	6.735 [ 1.643 ; 27.611 ]
atb.mssa.sensi.S	762	10.8	4.2	Z=2.94 d.f.=NA, P=0.0033	NA	3.162 [ 1.467 ; 6.814 ]
atb.hamrsa.sensi	762	LogRank Test	Overall effect	F=14.94 d.f.=3 P=0.0019		NA
atb.hamrsa.sensi.I	762	1.1	0.4	Z=0.9 d.f.=NA, P=0.3662	NA	2.843 [ 0.295 ; 27.425 ]
atb.hamrsa.sensi.R	762	12.9	4.6	Z=3.41 d.f.=NA, P=7e-04	NA	3.483 [ 1.7 ; 7.137 ]
atb.hamrsa.sensi.S	762	1.1	0	Z=0 d.f.=NA, P=0.9968	NA	178761151 [ 0 ; Inf ]
nb.nbgh.atb.expo	762	0/0.5/4	0/0.4/4	Z=0.99 d.f.=1 P=0.3231	0.76	1.172 [ 0.855 ; 1.607 ]
nb.nbgh.atb.mssa.sensi	762	0/0.2/3	0/0.2/3	Z=0.58 d.f.=1 P=0.5634	0.54	1.135 [ 0.738 ; 1.745 ]
nb.nbgh.atb.mrsa.sensi	762	0/0.1/2	0/0.1/2	Z=1.63 d.f.=1 P=0.1031	0.3	1.985 [ 0.87 ; 4.526 ]
previous.sa	762	LogRank Test	Overall effect	F=77.28 d.f.=1 P=0		NA
previous.sa.TRUE	762	76.3	29.6	Z=7.84 d.f.=NA, P=0	NA	8.288 [ 4.883 ; 14.066 ]
previous.sarm	762	LogRank Test	Overall effect	F=89.31 d.f.=1 P=0		NA
previous.sarm.TRUE	762	68.8	19.1	Z=8.64 d.f.=NA, P=0	NA	9.412 [ 5.661 ; 15.65 ]
previous.sasm	762	LogRank Test	Overall effect	F=7.45 d.f.=1 P=0.0063		NA
previous.sasm.TRUE	762	29	16.6	Z=2.83 d.f.=NA, P=0.0047	NA	2.102 [ 1.256 ; 3.517 ]
previous.same.strain	762	LogRank Test	Overall effect	F=73.47 d.f.=1 P=0		NA
previous.same.strain.TRUE	762	29	1.9	Z=7.89 d.f.=NA, P=0	NA	21.319 [ 9.974 ; 45.57 ]
prev.sa.nbgh	762	0/6.4/31	0/5.6/24	Z=2.26 d.f.=1 P=0.0235	4.59	1.067 [ 1.009 ; 1.129 ]
prev.sasm.nbgh	762	0/3.7/18	0/3.4/18	Z=1.51 d.f.=1 P=0.1319	2.98	1.065 [ 0.981 ; 1.156 ]
prev.sarm.nbgh	762	0/4.6/17	0/3.8/16	Z=2.75 d.f.=1 P=0.006	3.31	1.109 [ 1.03 ; 1.195 ]
prev.sa.nbgh.PA	762	0/3.3/21	0/2.9/20	Z=1.44 d.f.=1 P=0.1493	2.97	1.072 [ 0.975 ; 1.177 ]
prev.sasm.nbgh.PA	762	0/1.5/10	0/1.4/13	Z=0.52 d.f.=1 P=0.6033	1.66	1.041 [ 0.896 ; 1.209 ]
prev.sarm.nbgh.PA	762	0/2.4/14	0/2.1/13	Z=2 d.f.=1 P=0.0452	2.21	1.124 [ 1.003 ; 1.261 ]
prev.sa.nbgh.PE	762	0/3.2/18	0/2.7/14	Z=2.13 d.f.=1 P=0.033	2.69	1.091 [ 1.007 ; 1.182 ]
prev.sasm.nbgh.PE	762	0/2.2/12	0/2/11	Z=1.66 d.f.=1 P=0.0963	2.05	1.096 [ 0.984 ; 1.221 ]
prev.sarm.nbgh.PE	762	0/2.1/15	0/1.7/12	Z=2.29 d.f.=1 P=0.022	2.02	1.129 [ 1.018 ; 1.253 ]
prev.sa.nbgh.time	762	0/53.5/437.6	0/42.1/471	Z=1.78 d.f.=1 P=0.0752	72.61	1.003 [ 1 ; 1.005 ]
prev.sasm.nbgh.time	762	0/25.6/320.5	0/21.6/366.3	Z=1.2 d.f.=1 P=0.2317	44.72	1.003 [ 0.998 ; 1.007 ]
prev.sarm.nbgh.time	762	0/45.2/372.3	0/33.5/455.4	Z=1.83 d.f.=1 P=0.0667	67.9	1.003 [ 1 ; 1.006 ]

prev.sa.nbgh.PA.time	762	0/51.1/437.2	0/40.7/470.2	Z=1.63 d.f.=1 P=0.1024	72.54	1.002 [ 1 ; 1.005 ]
prev.sasm.nbgh.PA.time	762	0/24.1/320.1	0/20.7/365.5	Z=1.05 d.f.=1 P=0.2915	44.68	1.002 [ 0.998 ; 1.007 ]
prev.sarm.nbgh.PA.time	762	0/43.3/371.2	0/32.5/453.6	Z=1.7 d.f.=1 P=0.0897	67.84	1.003 [ 1 ; 1.005 ]
prev.sa.nbgh.PE.time	762	0/2.4/90.1	0/1.4/46.4	Z=1.86 d.f.=1 P=0.0634	4.84	1.035 [ 0.998 ; 1.072 ]
prev.sasm.nbgh.PE.time	762	0/1.6/46.9	0/1/30.6	Z=1.87 d.f.=1 P=0.0614	3.12	1.052 [ 0.998 ; 1.109 ]
prev.sarm.nbgh.PE.time	762	0/1.9/82.5	0/1/28.5	Z=1.83 d.f.=1 P=0.0678	4.16	1.041 [ 0.997 ; 1.088 ]
nw.range	762	1/10.1/57	1/9.1/42	Z=2.04 d.f.=1 P=0.0418	6.99	1.042 [ 1.002 ; 1.085 ]
betweenness.ego	762	0/9.8/100	0/8/70.8	Z=1.24 d.f.=1 P=0.2164	12.59	1.01 [ 0.994 ; 1.025 ]
mean.with.lvl1	762	0.5/3.7/9.7	0.5/3.2/11.2	Z=2.53 d.f.=1 P=0.0113	2.15	1.152 [ 1.033 ; 1.285 ]
mean.with.same.lvl	762	0/3.1/8.9	0/2.6/10.6	Z=2.59 d.f.=1 P=0.0096	2.15	1.154 [ 1.035 ; 1.285 ]
mean.at.lvl1	762	0.2/4/9.9	0.2/3.5/11.6	Z=2.59 d.f.=1 P=0.0097	2.21	1.15 [ 1.034 ; 1.278 ]
cpi.in.wk	762	1/19.9/63	1/17.5/71.5	Z=2.16 d.f.=1 P=0.0309	12.74	1.02 [ 1.002 ; 1.038 ]
max.dly.cpi	762	1/6.7/33	1/6.3/34	Z=1.37 d.f.=1 P=0.1709	5.45	1.039 [ 0.984 ; 1.097 ]
mean.dly.cpi	762	1/4.3/15	1/4.2/22	Z=0.86 d.f.=1 P=0.3892	3.28	1.034 [ 0.958 ; 1.117 ]
med.dly.cpi	762	1/4.1/14.5	1/3.9/22	Z=0.94 d.f.=1 P=0.3469	3.2	1.036 [ 0.962 ; 1.115 ]
max.dur.dly.cpi	762	0.1/12.2/24	0.1/11.7/24	Z=0.82 d.f.=1 P=0.4113	7.71	1.012 [ 0.983 ; 1.043 ]
mean.dur.dly.cpi	762	0.1/3.6/13.6	0.1/3.6/13.4	Z=0.21 d.f.=1 P=0.8311	2.78	1.009 [ 0.93 ; 1.094 ]
med.dur.dly.cpi	762	0.1/2/15.6	0.1/2/13.4	Z=0.33 d.f.=1 P=0.7404	2.64	1.014 [ 0.933 ; 1.103 ]
degree	762	1/10.1/57	1/9.1/42	Z=2.04 d.f.=1 P=0.0418	6.99	1.042 [ 1.002 ; 1.085 ]
degree.PA	762	0/5.7/39	0/5.3/30	Z=1.03 d.f.=1 P=0.3022	4.83	1.031 [ 0.973 ; 1.093 ]
degree.PE	762	0/4.5/21	0/3.8/17	Z=2.27 d.f.=1 P=0.0231	3.57	1.076 [ 1.01 ; 1.147 ]
degree.PA.sa	762	0/1.9/11	0/1.8/10	Z=0.65 d.f.=1 P=0.5129	1.95	1.044 [ 0.918 ; 1.186 ]
degree.PA.sasm	762	0/0.7/5	0/0.8/5	Z=-1.23 d.f.=1 P=0.2183	1	0.848 [ 0.652 ; 1.103 ]
degree.PA.sarm	762	0/1.4/9	0/1.1/8	Z=2 d.f.=1 P=0.0455	1.45	1.178 [ 1.003 ; 1.383 ]
degree.PE.sa	762	0/1.4/10	0/1/5	Z=3.34 d.f.=1 P=8e-04	1.28	1.308 [ 1.117 ; 1.531 ]
degree.PE.sasm	762	0/1/8	0/0.7/5	Z=3.52 d.f.=1 P=4e-04	1.03	1.416 [ 1.167 ; 1.72 ]
degree.PE.sarm	762	0/0.5/3	0/0.4/5	Z=1.75 d.f.=1 P=0.0802	0.69	1.306 [ 0.968 ; 1.762 ]
degree.sameWard	762	0/6.9/18	0/6.1/19	Z=2.14 d.f.=1 P=0.0322	4.16	1.059 [ 1.005 ; 1.116 ]
degree.otherWard	762	0/3.2/49	0/3/35	Z=0.65 d.f.=1 P=0.5132	5.35	1.018 [ 0.965 ; 1.073 ]
degree.PA.sameWard	762	0/3.9/13	0/3.7/14	Z=0.5 d.f.=1 P=0.6191	2.84	1.02 [ 0.943 ; 1.104 ]
degree.PE.sameWard	762	0/3.1/17	0/2.4/14	Z=2.91 d.f.=1 P=0.0037	2.58	1.125 [ 1.039 ; 1.218 ]
degree.PA.otherWard	762	0/1.8/35	0/1.6/21	Z=0.92 d.f.=1 P=0.3594	3.7	1.036 [ 0.961 ; 1.117 ]
degree.PE.otherWard	762	0/1.4/14	0/1.4/14	Z=0.04 d.f.=1 P=0.9703	2.11	1.002 [ 0.886 ; 1.134 ]
degree.PE.nightshift	762	0/0.8/4	0/0.7/6	Z=1.59 d.f.=1 P=0.1123	0.97	1.197 [ 0.959 ; 1.495 ]
degree.PE.ergo	762	0/0/1	0/0.1/2	Z=-0.54 d.f.=1 P=0.5891	0.28	0.786 [ 0.327 ; 1.886 ]
degree.PE.kine	762	0/0.2/4	0/0.2/5	Z=-0.92 d.f.=1 P=0.359	0.58	0.785 [ 0.467 ; 1.317 ]
degree.PE.other	762	0/0.1/3	0/0.1/6	Z=-0.36 d.f.=1 P=0.7156	0.5	0.892 [ 0.484 ; 1.646 ]
nbgh.of.nbgh.bw	762	0/12.9/76	0/12.8/59.9	Z=-0.16 d.f.=1 P=0.8761	8	0.998 [ 0.97 ; 1.026 ]
nbgh.of.nbgh.nb	762	1/11.6/30.6	1/11/34.6	Z=2.07 d.f.=1 P=0.0383	5.25	1.071 [ 1.004 ; 1.143 ]
nbgh.of.nbgh.nb.PA	762	0/6.8/20.5	1/6.6/22.9	Z=0.86 d.f.=1 P=0.391	3.61	1.042 [ 0.949 ; 1.143 ]
nbgh.of.nbgh.nb.PE	762	0/4.9/14.3	0/4.4/13.2	Z=2.53 d.f.=1 P=0.0113	2.57	1.139 [ 1.03 ; 1.259 ]
nbgh.of.nbgh.cpi.nb	762	1/19.5/39	1/18/44.1	Z=2.53 d.f.=1 P=0.0113	8.16	1.044 [ 1.01 ; 1.078 ]
nbgh.of.nbgh.cpi.dur	762	0.1/51/395.3	0.4/43.6/257.7	Z=1.85 d.f.=1 P=0.0639	45.25	1.004 [ 1 ; 1.008 ]
nbgh.of.nbgh.cpi.dur.PA	762	0/48.4/394.8	0.1/41.2/257.1	Z=1.77 d.f.=1 P=0.0765	45.65	1.004 [ 1 ; 1.008 ]
nbgh.of.nbgh.cpi.dur.PE	762	0/2.6/13.9	0/2.3/14.3	Z=1.37 d.f.=1 P=0.172	2.41	1.066 [ 0.972 ; 1.169 ]
nbgh.of.nbgh.cpi.dly.nb	762	2/9.8/28.8	2/9.6/31.9	Z=0.91 d.f.=1 P=0.3607	4.88	1.028 [ 0.969 ; 1.091 ]
nbgh.of.nbgh.cpi.dly.dur	762	0.1/2.3/10.1	0.1/2.3/12.2	Z=0.59 d.f.=1 P=0.5527	1.57	1.042 [ 0.911 ; 1.191 ]
dens	762	17.6/54.1/100	19.4/54/100	Z=-0.28 d.f.=1 P=0.7802	20.25	0.998 [ 0.987 ; 1.01 ]
dens.pat	762	0/54.8/100	0/60.7/100	Z=-2.24 d.f.=1 P=0.0248	26.77	0.991 [ 0.982 ; 0.999 ]
dur.cont	762	0.1/78.4/472.8	0.1/69.7/539.8	Z=1.28 d.f.=1 P=0.1994	89.7	1.002 [ 0.999 ; 1.004 ]

dur.cont.PA	762	0/75.2/470.8	0/67.8/536.8	Z=1.13 d.f.=1 P=0.2596	89.77	1.001 [ 0.999 ; 1.004 ]
dur.cont.PE	762	0/3.2/90.2	0/1.9/68.6	Z=2.13 d.f.=1 P=0.0331	5.62	1.034 [ 1.003 ; 1.066 ]
dur.cont.uncolonized	762	0.1/44/437.6	0/49/539.1	Z=-0.29 d.f.=1 P=0.7697	71.36	0.999 [ 0.996 ; 1.003 ]
dur.cont.uncolonized.PA	762	0/41.9/437.2	0/47.6/536.8	Z=-0.4 d.f.=1 P=0.6861	71.36	0.999 [ 0.996 ; 1.003 ]
dur.cont.uncolonized.PE	762	0/2.1/70.2	0/1.4/57.1	Z=1.6 d.f.=1 P=0.1099	4.32	1.032 [ 0.993 ; 1.073 ]
dur.cont.no.sasm	762	0.1/66.2/437.6	0/59.4/539.4	Z=1.23 d.f.=1 P=0.2175	79.82	1.002 [ 0.999 ; 1.004 ]
dur.cont.no.sasm.PA	762	0/63.7/437.2	0/57.8/536.8	Z=1.11 d.f.=1 P=0.2681	79.8	1.002 [ 0.999 ; 1.004 ]
dur.cont.no.sasm.PE	762	0/2.5/70.2	0/1.5/57.1	Z=1.96 d.f.=1 P=0.05	4.56	1.038 [ 1 ; 1.078 ]
dur.cont.no.sarm	762	0.1/54.2/437.6	0/58.4/539.5	Z=-0.17 d.f.=1 P=0.8642	79.02	1 [ 0.997 ; 1.003 ]
dur.cont.no.sarm.PA	762	0/51.5/437.2	0/56.7/536.8	Z=-0.32 d.f.=1 P=0.7526	79.08	1 [ 0.996 ; 1.003 ]
dur.cont.no.sarm.PE	762	0/2.7/89	0/1.7/68.6	Z=1.83 d.f.=1 P=0.0671	5.27	1.03 [ 0.998 ; 1.063 ]
rel.time.PA	762	0/85.7/100	0/92.2/100	Z=-3.07 d.f.=1 P=0.0021	19.24	0.986 [ 0.977 ; 0.995 ]
rel.time.PE	762	0/14.3/100	0/7.8/100	Z=3.07 d.f.=1 P=0.0021	19.24	1.014 [ 1.005 ; 1.024 ]
rel.time.sa	762	0/39.3/99.4	0/29.1/100	Z=2.74 d.f.=1 P=0.0062	34.09	1.009 [ 1.002 ; 1.015 ]
rel.time.uncolo	762	0.6/60.7/100	0/70.9/100	Z=-2.74 d.f.=1 P=0.0062	34.09	0.991 [ 0.985 ; 0.998 ]
rel.time.PA.sa	762	0/33.8/99.4	0/27.1/100	Z=1.81 d.f.=1 P=0.0707	33.45	1.006 [ 1 ; 1.012 ]
rel.time.PA.uncolo	762	0/51.9/100	0/65.2/100	Z=-3.3 d.f.=1 P=0.001	36.59	0.99 [ 0.984 ; 0.996 ]
rel.time.PE.sa	762	0/5.5/78.7	0/2/100	Z=3.07 d.f.=1 P=0.0021	9.3	1.026 [ 1.009 ; 1.043 ]
rel.time.PE.uncolo	762	0/8.8/100	0/5.8/100	Z=1.88 d.f.=1 P=0.06	14.79	1.012 [ 0.999 ; 1.025 ]
rel.time.sasm	762	0/11.8/97.6	0/14.9/100	Z=-1.42 d.f.=1 P=0.1568	24.24	0.992 [ 0.982 ; 1.003 ]
rel.time.no.sasm	762	2.4/88.2/100	0/85.1/100	Z=1.42 d.f.=1 P=0.1568	24.24	1.008 [ 0.997 ; 1.018 ]
rel.time.PA.sasm	762	0/9.6/97.6	0/13.3/99.7	Z=-1.76 d.f.=1 P=0.0791	23.52	0.99 [ 0.978 ; 1.001 ]
rel.time.PA.no.sasm	762	0/76.2/100	0/78.9/100	Z=-0.67 d.f.=1 P=0.5031	28.71	0.997 [ 0.99 ; 1.005 ]
rel.time.PE.sasm	762	0/2.3/66.2	0/1.6/100	Z=0.95 d.f.=1 P=0.3415	7.52	1.011 [ 0.988 ; 1.035 ]
rel.time.PE.no.sasm	762	0/12/100	0/6.2/100	Z=3.12 d.f.=1 P=0.0018	16.34	1.017 [ 1.006 ; 1.028 ]
rel.time.sarm	762	0/30/98.8	0/15.4/100	Z=4.83 d.f.=1 P=0	28.45	1.016 [ 1.01 ; 1.023 ]
rel.time.no.sarm	762	1.2/70/100	0/84.6/100	Z=-4.83 d.f.=1 P=0	28.45	0.984 [ 0.977 ; 0.99 ]
rel.time.PA.sarm	762	0/26.6/98.8	0/14.6/100	Z=4.16 d.f.=1 P=0	27.75	1.015 [ 1.008 ; 1.022 ]
rel.time.PA.no.sarm	762	0/59.1/100	0/77.6/100	Z=-5.23 d.f.=1 P=0	33.01	0.984 [ 0.978 ; 0.99 ]
rel.time.PE.sarm	762	0/3.4/78.7	0/0.8/100	Z=2.65 d.f.=1 P=0.0079	7.35	1.028 [ 1.007 ; 1.049 ]
rel.time.PE.no.sarm	762	0/10.8/100	0/7/100	Z=2.14 d.f.=1 P=0.0325	17.06	1.012 [ 1.001 ; 1.023 ]
nb.sa	762	0/3.7/17	0/2.8/11	Z=3.89 d.f.=1 P=1e-04	2.48	1.23 [ 1.108 ; 1.365 ]
dens.sa	762	0/32.5/88.9	0/26.6/80	Z=3.18 d.f.=1 P=0.0015	18.44	1.021 [ 1.008 ; 1.034 ]
dur.sa	762	0/34.5/341.9	0/20.7/307.4	Z=2.83 d.f.=1 P=0.0047	44.9	1.006 [ 1.002 ; 1.01 ]
dur.PA.sa	762	0/33.3/340.3	0/20.2/306.9	Z=2.69 d.f.=1 P=0.007	44.92	1.005 [ 1.001 ; 1.009 ]
dur.PE.sa	762	0/1.1/30.2	0/0.5/11.5	Z=2.77 d.f.=1 P=0.0057	1.82	1.153 [ 1.042 ; 1.275 ]
nb.sarm	762	0/2.1/10	0/1.5/8	Z=4.07 d.f.=1 P=0	1.73	1.345 [ 1.166 ; 1.551 ]
dens.sarm	762	0/19/57.1	0/13.6/60	Z=3.77 d.f.=1 P=2e-04	14.34	1.031 [ 1.015 ; 1.048 ]
dur.sarm	762	0/24.2/259.4	0/11.2/210.3	Z=3.66 d.f.=1 P=3e-04	31.82	1.01 [ 1.005 ; 1.016 ]
dur.PA.sarm	762	0/23.7/258.4	0/11.1/210.3	Z=3.58 d.f.=1 P=3e-04	31.8	1.01 [ 1.004 ; 1.015 ]
dur.PE.sarm	762	0/0.5/12.6	0/0.2/6.7	Z=2.84 d.f.=1 P=0.0045	0.89	1.301 [ 1.085 ; 1.561 ]
nb.sasm	762	0/1.8/9	0/1.5/7	Z=2.41 d.f.=1 P=0.0157	1.43	1.228 [ 1.039 ; 1.45 ]
dens.sasm	762	0/16.2/50	0/14.8/75	Z=1.1 d.f.=1 P=0.2694	13.16	1.01 [ 0.993 ; 1.027 ]
dur.sasm	762	0/12.2/124.6	0/10.3/256.4	Z=0.68 d.f.=1 P=0.4973	26.52	1.003 [ 0.995 ; 1.01 ]
dur.PA.sasm	762	0/11.5/124.5	0/9.9/256.4	Z=0.53 d.f.=1 P=0.5979	26.53	1.002 [ 0.994 ; 1.01 ]
dur.PE.sasm	762	0/0.7/28.6	0/0.3/11.5	Z=2.1 d.f.=1 P=0.0359	1.53	1.126 [ 1.008 ; 1.257 ]
prev.same.strain.nbgh	762	0/0.7/4	0/0.3/7	Z=4.25 d.f.=1 P=0	0.86	1.706 [ 1.334 ; 2.181 ]
dens.same.strain	762	0/3.4/42.9	0/1/100	Z=2.61 d.f.=1 P=0.0092	5.73	1.056 [ 1.014 ; 1.1 ]
nb.same.strain	762	0/0.3/3	0/0.1/4	Z=3.06 d.f.=1 P=0.0022	0.43	2.032 [ 1.291 ; 3.198 ]
PA.same.strain	762	0/0.2/3	0/0.1/2	Z=3.38 d.f.=1 P=7e-04	0.32	2.668 [ 1.511 ; 4.71 ]

PE.same.strain	762	0/0.1/1	0/0/3	Z=0.65 d.f.=1 P=0.5173	0.24	1.315 [ 0.574 ; 3.012 ]	
rel.dur.same.strain	762	0/6/98.8	0/0.8/100	Z=3.22 d.f.=1 P=0.0013	9.3	1.047 [ 1.018 ; 1.077 ]	
rel.dur.PA.same.strain	762	0/5.9/98.3	0/0.8/100	Z=3.22 d.f.=1 P=0.0013	9.28	1.048 [ 1.018 ; 1.078 ]	
rel.dur.PE.same.strain	762	0/0/0.5	0/0.1/7.2	Z=-0.67 d.f.=1 P=0.5003	0.47	0.637 [ 0.171 ; 2.366 ]	
dur.cont.same.strain	762	0/3.5/58.4	0/0.9/110.1	Z=2.36 d.f.=1 P=0.0185	8.04	1.025 [ 1.004 ; 1.046 ]	
dur.cont.PA.same.strain	762	0/3.4/58.1	0/0.8/110.1	Z=2.36 d.f.=1 P=0.0183	8.03	1.025 [ 1.004 ; 1.046 ]	
dur.cont.PE.same.strain	762	0/0/1	0/0/3.8	Z=-0.17 d.f.=1 P=0.8625	0.22	0.901 [ 0.277 ; 2.929 ]	
mean.same.strain.nbgh	762	0/0.2/2	0/0.1/2.8	Z=3.63 d.f.=1 P=3e-04	0.33	4.808 [ 2.059 ; 11.229 ]	
mean.same.strain.nbgh.PA	762	0/0.1/2	0/0.1/2.4	Z=2.45 d.f.=1 P=0.0144	0.24	3.66 [ 1.295 ; 10.341 ]	
mean.same.strain.nbgh.PE	762	0/0.1/0.8	0/0.1/1.7	Z=3.28 d.f.=1 P=0.001	0.17	8.471 [ 2.36 ; 30.403 ]	
mean.sarm.2hop.nbgh	762	0/2/5.2	0/1.7/5.7	Z=3.46 d.f.=1 P=5e-04	1.25	1.527 [ 1.201 ; 1.941 ]	
mean.sarm.2hop.nbgh.viaPA	762	0/0.9/3.6	0/0.9/4.7	Z=0.23 d.f.=1 P=0.8165	0.96	1.035 [ 0.776 ; 1.378 ]	
mean.sarm.2hop.nbgh.viaPE	762	0/1.1/4	0/0.7/4.7	Z=4.02 d.f.=1 P=1e-04	0.82	1.773 [ 1.341 ; 2.345 ]	



## Annexe 2 : Résultats de l'analyse cas-témoins pour la colonisation des patients par SASM

Un dictionnaire des covariables est proposé en Annexe 3.

Univariate - SA_type = SASM - Level = 1 - Lag = 7 - Duration = >300						
	N	CASE (N=67 ) min/mean/max	CTRL (N=526 ) min/mean/max	Test Statistic	SD	Odds Ratio per unit
age	593	32/60.1/91	27/59.6/103	Z=0.19 d.f.=1 P=0.8485	17.1	1.001 [ 0.986 ; 1.017 ]
sexe	593	LogRank Test	Overall effect	F=3.3 d.f.=1 P=0.0691		NA
sexe.M	593	58.2	46.4	Z=1.81 d.f.=NA, P=0.071	NA	1.596 [ 0.961 ; 2.65 ]
mif	593	LogRank Test	Overall effect	F=5.3 d.f.=3 P=0.1514		NA
mif.[21,48]	593	32.8	24.7	Z=0.98 d.f.=NA, P=0.3289	NA	1.403 [ 0.711 ; 2.767 ]
mif.(48,85]	593	14.9	26	Z=-1.22 d.f.=NA, P=0.2216	NA	0.599 [ 0.263 ; 1.363 ]
mif.(85,124]	593	28.4	24.1	Z=0.56 d.f.=NA, P=0.5788	NA	1.219 [ 0.605 ; 2.456 ]
time.at.risk	590	5/53.4/133	4/52.7/152	Z=0.13 d.f.=1 P=0.8935	42.13	1 [ 0.994 ; 1.006 ]
atb.expo	593	LogRank Test	Overall effect	F=0.83 d.f.=1 P=0.3628		NA
atb.expo.TRUE	593	7.5	4.8	Z=0.96 d.f.=NA, P=0.3391	NA	1.616 [ 0.604 ; 4.324 ]
atb.mssa.sensi	593	LogRank Test	Overall effect	F=1.57 d.f.=2 P=0.4569		NA
atb.mssa.sensi.R	593	1.5	1.3	Z=0.12 d.f.=NA, P=0.9018	NA	1.141 [ 0.14 ; 9.281 ]
atb.mssa.sensi.S	593	6	2.9	Z=1.34 d.f.=NA, P=0.1792	NA	2.149 [ 0.704 ; 6.56 ]
atb.hamrsa.sensi	593	LogRank Test	Overall effect	F=6.99 d.f.=2 P=0.0304		NA
atb.hamrsa.sensi.I	593	4.5	0.4	Z=2.71 d.f.=NA, P=0.0067	NA	11.901 [ 1.986 ; 71.3 ]
atb.hamrsa.sensi.R	593	3	3.8	Z=-0.24 d.f.=NA, P=0.8089	NA	0.835 [ 0.195 ; 3.585 ]
nb.nbgh.atb.expo	593	0/0.3/2	0/0.3/4	Z=-0.75 d.f.=1 P=0.4506	0.63	0.838 [ 0.53 ; 1.326 ]
nb.nbgh.atb.mssa.sensi	593	0/0.2/2	0/0.2/2	Z=-0.03 d.f.=1 P=0.9724	0.47	0.99 [ 0.575 ; 1.707 ]
nb.nbgh.atb.mrsa.sensi	593	0/0/1	0/0.1/1	Z=-0.39 d.f.=1 P=0.6956	0.23	0.779 [ 0.224 ; 2.715 ]
previous.sa	593	LogRank Test	Overall effect	F=62.69 d.f.=1 P=0		NA
previous.sa.TRUE	593	79.1	28.5	Z=7 d.f.=NA, P=0	NA	8.922 [ 4.833 ; 16.468 ]
previous.sarm	593	LogRank Test	Overall effect	F=8.5 d.f.=1 P=0.0035		NA
previous.sarm.TRUE	593	34.3	18.3	Z=3.02 d.f.=NA, P=0.0025	NA	2.34 [ 1.348 ; 4.061 ]
previous.sasm	593	LogRank Test	Overall effect	F=78.19 d.f.=1 P=0		NA
previous.sasm.TRUE	593	70.1	16	Z=8.19 d.f.=NA, P=0	NA	11.051 [ 6.221 ; 19.634 ]
previous.same.strain	593	LogRank Test	Overall effect	F=107.08 d.f.=1 P=0		NA
previous.same.strain.TRUE	593	40.3	0.6	Z=5.23 d.f.=NA, P=0	NA	205.208 [ 27.861 ; 1511.467 ]
prev.sa.nbgh	593	0/4.4/12	0/4.9/25	Z=-1.03 d.f.=1 P=0.3037	3.59	0.959 [ 0.887 ; 1.038 ]
prev.sasm.nbgh	593	0/2.8/9	0/3/18	Z=-0.68 d.f.=1 P=0.4937	2.6	0.963 [ 0.866 ; 1.072 ]
prev.sarm.nbgh	593	0/2.9/9	0/3.2/16	Z=-1.08 d.f.=1 P=0.2822	2.66	0.943 [ 0.848 ; 1.049 ]
prev.sa.nbgh.PA	593	0/2/6	0/2.3/11	Z=-1.23 d.f.=1 P=0.2199	2.11	0.918 [ 0.8 ; 1.053 ]
prev.sasm.nbgh.PA	593	0/1/4	0/1.2/6	Z=-0.74 d.f.=1 P=0.4621	1.3	0.924 [ 0.749 ; 1.141 ]
prev.sarm.nbgh.PA	593	0/1.3/6	0/1.6/10	Z=-1.35 d.f.=1 P=0.1784	1.73	0.889 [ 0.748 ; 1.055 ]
prev.sa.nbgh.PE	593	0/2.4/10	0/2.5/19	Z=-0.46 d.f.=1 P=0.6444	2.43	0.974 [ 0.872 ; 1.088 ]
prev.sasm.nbgh.PE	593	0/1.7/8	0/1.8/14	Z=-0.41 d.f.=1 P=0.6793	1.97	0.972 [ 0.848 ; 1.113 ]
prev.sarm.nbgh.PE	593	0/1.5/8	0/1.6/13	Z=-0.3 d.f.=1 P=0.7617	1.71	0.975 [ 0.829 ; 1.147 ]
prev.sa.nbgh.time	593	0/40.1/389.7	0/38.1/555.5	Z=0.25 d.f.=1 P=0.8027	76.45	1 [ 0.997 ; 1.004 ]
prev.sasm.nbgh.time	593	0/23.6/190.9	0/21.4/434.3	Z=0.39 d.f.=1 P=0.6987	49.95	1.001 [ 0.996 ; 1.006 ]
prev.sarm.nbgh.time	593	0/29.9/389.5	0/29/473.7	Z=0.14 d.f.=1 P=0.8848	70.39	1 [ 0.997 ; 1.004 ]
prev.sa.nbgh.PA.time	593	0/39.1/389.2	0/36.9/554	Z=0.27 d.f.=1 P=0.7887	76.44	1 [ 0.997 ; 1.004 ]

prev.sasm.nbgh.PA.time	593	0/23/190.7	0/20.5/432.9	Z=0.43 d.f.=1 P=0.6688	49.94	1.001 [ 0.996 ; 1.006 ]
prev.sarm.nbgh.PA.time	593	0/29.1/389.2	0/28.2/473.4	Z=0.14 d.f.=1 P=0.8879	70.35	1 [ 0.997 ; 1.004 ]
prev.sa.nbgh.PE.time	593	0/1.1/10.5	0/1.2/27.9	Z=-0.48 d.f.=1 P=0.6308	2.99	0.976 [ 0.882 ; 1.079 ]
prev.sasm.nbgh.PE.time	593	0/0.6/5.1	0/0.9/21.4	Z=-0.88 d.f.=1 P=0.3766	2.28	0.93 [ 0.792 ; 1.092 ]
prev.sarm.nbgh.PE.time	593	0/0.8/10.5	0/0.8/21.7	Z=0.13 d.f.=1 P=0.8954	2.31	1.008 [ 0.898 ; 1.132 ]
nw.range	593	1/7.8/25	1/8.2/41	Z=-0.54 d.f.=1 P=0.5871	5.31	0.986 [ 0.935 ; 1.039 ]
betweenness.ego	593	0/8.4/100	0/9/100	Z=-0.34 d.f.=1 P=0.7368	14.92	0.997 [ 0.979 ; 1.015 ]
mean.with.lvl1	593	0.5/3/7.5	0.5/3/14.6	Z=0.31 d.f.=1 P=0.7547	1.96	1.022 [ 0.893 ; 1.17 ]
mean.with.same.lvl	593	0/2.4/7.3	0/2.4/14	Z=0.41 d.f.=1 P=0.6799	1.98	1.028 [ 0.9 ; 1.174 ]
mean.at.lvl1	593	0.2/3.4/8.3	0.2/3.3/15	Z=0.45 d.f.=1 P=0.6506	2.03	1.03 [ 0.905 ; 1.173 ]
cpi.in.wk	593	1/17/58.5	1/16.6/64	Z=0.38 d.f.=1 P=0.7005	12.58	1.004 [ 0.983 ; 1.026 ]
max.dly.cpi	593	1/4.9/14	1/5.3/25	Z=-0.68 d.f.=1 P=0.4949	3.52	0.972 [ 0.895 ; 1.055 ]
mean.dly.cpi	593	1/3.6/14	1/3.6/19	Z=-0.11 d.f.=1 P=0.9134	2.49	0.994 [ 0.889 ; 1.111 ]
med.dly.cpi	593	1/3.5/14	1/3.5/19	Z=0.12 d.f.=1 P=0.9012	2.51	1.007 [ 0.903 ; 1.122 ]
max.dur.dly.cpi	593	0.1/12.3/24	0.1/11.1/24	Z=1.34 d.f.=1 P=0.1809	7.69	1.023 [ 0.99 ; 1.056 ]
mean.dur.dly.cpi	593	0.1/3.8/15.5	0.1/3.4/11.8	Z=1.21 d.f.=1 P=0.2267	2.76	1.056 [ 0.967 ; 1.153 ]
med.dur.dly.cpi	593	0.1/2.2/14.5	0.1/2/11.3	Z=0.55 d.f.=1 P=0.5808	2.62	1.027 [ 0.935 ; 1.128 ]
degree	593	1/7.8/25	1/8.2/41	Z=-0.54 d.f.=1 P=0.5871	5.31	0.986 [ 0.935 ; 1.039 ]
degree.PA	593	0/4.3/16	0/4.5/20	Z=-0.31 d.f.=1 P=0.7535	3.41	0.987 [ 0.908 ; 1.072 ]
degree.PE	593	0/3.5/12	0/3.7/26	Z=-0.53 d.f.=1 P=0.5929	3.34	0.978 [ 0.901 ; 1.061 ]
degree.PA.sa	593	0/1.4/5	0/1.5/9	Z=-0.39 d.f.=1 P=0.6983	1.62	0.966 [ 0.813 ; 1.149 ]
degree.PA.sasm	593	0/0.6/3	0/0.6/4	Z=-0.1 d.f.=1 P=0.9198	0.89	0.984 [ 0.726 ; 1.335 ]
degree.PA.sarm	593	0/0.8/4	0/0.9/8	Z=-0.52 d.f.=1 P=0.5998	1.15	0.937 [ 0.734 ; 1.196 ]
degree.PE.sa	593	0/0.8/8	0/0.8/6	Z=-0.02 d.f.=1 P=0.9835	1.16	0.998 [ 0.791 ; 1.258 ]
degree.PE.sasm	593	0/0.5/5	0/0.6/6	Z=-0.68 d.f.=1 P=0.4956	0.94	0.902 [ 0.67 ; 1.214 ]
degree.PE.sarm	593	0/0.3/3	0/0.3/3	Z=1.19 d.f.=1 P=0.2351	0.55	1.326 [ 0.832 ; 2.112 ]
degree.sameWard	593	0/5.9/15	0/6/19	Z=-0.17 d.f.=1 P=0.8656	3.78	0.994 [ 0.926 ; 1.066 ]
degree.otherWard	593	0/1.9/15	0/2.2/27	Z=-0.67 d.f.=1 P=0.5015	3.29	0.969 [ 0.885 ; 1.062 ]
degree.PA.sameWard	593	0/3.5/12	0/3.5/14	Z=0.01 d.f.=1 P=0.9903	2.53	1.001 [ 0.9 ; 1.112 ]
degree.PE.sameWard	593	0/2.4/10	0/2.5/15	Z=-0.26 d.f.=1 P=0.7924	2.5	0.986 [ 0.887 ; 1.096 ]
degree.PA.otherWard	593	0/0.8/11	0/1/17	Z=-0.5 d.f.=1 P=0.6172	2.16	0.966 [ 0.843 ; 1.107 ]
degree.PE.otherWard	593	0/1.1/6	0/1.2/11	Z=-0.7 d.f.=1 P=0.4852	1.62	0.939 [ 0.787 ; 1.12 ]
degree.PE.nightshift	593	0/0.7/4	0/0.7/6	Z=0.24 d.f.=1 P=0.8083	0.99	1.033 [ 0.795 ; 1.343 ]
degree.PE.ergo	593	0/0.1/1	0/0.1/2	Z=-0.18 d.f.=1 P=0.8547	0.28	0.915 [ 0.355 ; 2.358 ]
degree.PE.kine	593	0/0.1/1	0/0.1/5	Z=-1.29 d.f.=1 P=0.1954	0.5	0.526 [ 0.199 ; 1.391 ]
degree.PE.other	593	0/0/1	0/0.1/2	Z=-1.46 d.f.=1 P=0.1455	0.25	0.229 [ 0.031 ; 1.667 ]
nbgh.of.nbgh.bw	593	0/12.4/35.6	0/14/100	Z=-1.49 d.f.=1 P=0.1356	9.09	0.973 [ 0.938 ; 1.009 ]
nbgh.of.nbgh.nb	593	1/10.1/18.2	1/10.5/24.4	Z=-0.86 d.f.=1 P=0.3908	3.86	0.964 [ 0.887 ; 1.048 ]
nbgh.of.nbgh.nb.PA	593	1/6/12.2	0/6.2/14.3	Z=-0.25 d.f.=1 P=0.8027	2.67	0.985 [ 0.872 ; 1.112 ]
nbgh.of.nbgh.nb.PE	593	0/4.1/10.6	0/4.3/14	Z=-0.96 d.f.=1 P=0.3393	2.35	0.94 [ 0.828 ; 1.067 ]
nbgh.of.nbgh.cpi.nb	593	1/18/41.7	3/17.6/42.1	Z=0.6 d.f.=1 P=0.5497	7.41	1.013 [ 0.971 ; 1.056 ]
nbgh.of.nbgh.cpi.dur	593	0.1/49.6/267.9	0.4/40.6/230	Z=1.76 d.f.=1 P=0.0788	41.44	1.005 [ 0.999 ; 1.01 ]
nbgh.of.nbgh.cpi.dur.PA	593	0.1/47.1/266	0/38.4/228.1	Z=1.71 d.f.=1 P=0.0864	41.79	1.005 [ 0.999 ; 1.01 ]
nbgh.of.nbgh.cpi.dur.PE	593	0/2.4/15.4	0/2.2/12.3	Z=0.57 d.f.=1 P=0.5674	2.17	1.034 [ 0.921 ; 1.161 ]
nbgh.of.nbgh.cpi.dly.nb	593	2/8.6/22.7	2/9/27.3	Z=-0.98 d.f.=1 P=0.3253	3.58	0.957 [ 0.878 ; 1.044 ]
nbgh.of.nbgh.cpi.dly.dur	593	0.1/2.4/11.4	0.1/2.1/7.8	Z=1.82 d.f.=1 P=0.0685	1.4	1.174 [ 0.988 ; 1.394 ]
dens	593	20.3/56.4/100	22.2/54.3/100	Z=0.77 d.f.=1 P=0.4416	18.98	1.005 [ 0.992 ; 1.018 ]
dens.pat	593	0/61.7/100	0/56.4/100	Z=1.66 d.f.=1 P=0.0966	27.09	1.008 [ 0.999 ; 1.018 ]
dur.cont	593	0.1/74.5/584.7	0.1/65.4/568.7	Z=0.82 d.f.=1 P=0.4097	93.6	1.001 [ 0.999 ; 1.004 ]
dur.cont.PA	593	0/72.5/583.8	0/63.7/567.7	Z=0.8 d.f.=1 P=0.4223	93.58	1.001 [ 0.999 ; 1.004 ]

dur.cont.PE	593	0/1.9/34.8	0/1.6/31.4	Z=0.57 d.f.=1 P=0.569	3.81	1.019 [ 0.956 ; 1.085 ]
dur.cont.uncolonized	593	0/51.7/460.2	0/46.3/568.1	Z=0.68 d.f.=1 P=0.4977	69.92	1.001 [ 0.998 ; 1.005 ]
dur.cont.uncolonized.PA	593	0/50.1/459.2	0/45/567.7	Z=0.64 d.f.=1 P=0.5219	69.94	1.001 [ 0.998 ; 1.005 ]
dur.cont.uncolonized.PE	593	0/1.6/34.8	0/1.2/26.1	Z=0.79 d.f.=1 P=0.4299	3.29	1.028 [ 0.96 ; 1.101 ]
dur.cont.no.sasm	593	0/65.2/460.2	0/57/568.1	Z=0.83 d.f.=1 P=0.4092	84.77	1.001 [ 0.998 ; 1.004 ]
dur.cont.no.sasm.PA	593	0/63.5/459.2	0/55.6/567.7	Z=0.79 d.f.=1 P=0.4269	84.73	1.001 [ 0.998 ; 1.004 ]
dur.cont.no.sasm.PE	593	0/1.7/34.8	0/1.3/26.4	Z=0.79 d.f.=1 P=0.4295	3.37	1.028 [ 0.96 ; 1.1 ]
dur.cont.no.sarm	593	0.1/60.9/584.7	0/54.5/568.4	Z=0.71 d.f.=1 P=0.4771	78.52	1.001 [ 0.998 ; 1.004 ]
dur.cont.no.sarm.PA	593	0/59.1/583.8	0/53/567.7	Z=0.68 d.f.=1 P=0.4937	78.59	1.001 [ 0.998 ; 1.004 ]
dur.cont.no.sarm.PE	593	0/1.8/34.8	0/1.5/31.2	Z=0.57 d.f.=1 P=0.5675	3.7	1.019 [ 0.956 ; 1.086 ]
rel.time.PA	593	0/93.3/100	0/90/100	Z=1.37 d.f.=1 P=0.1703	21.77	1.011 [ 0.995 ; 1.027 ]
rel.time.PE	593	0/6.7/100	0/10/100	Z=-1.37 d.f.=1 P=0.1703	21.77	0.989 [ 0.974 ; 1.005 ]
rel.time.sa	593	0/29.1/100	0/27.1/100	Z=0.49 d.f.=1 P=0.6229	33.36	1.002 [ 0.994 ; 1.009 ]
rel.time.uncolo	593	0/70.9/100	0/72.9/100	Z=-0.49 d.f.=1 P=0.6229	33.36	0.998 [ 0.991 ; 1.006 ]
rel.time.PA.sa	593	0/28.1/100	0/25/100	Z=0.81 d.f.=1 P=0.4197	32.6	1.003 [ 0.995 ; 1.011 ]
rel.time.PA.uncolo	593	0/65.1/100	0/65/100	Z=0.11 d.f.=1 P=0.9139	36.61	1 [ 0.993 ; 1.007 ]
rel.time.PE.sa	593	0/0.9/16.7	0/2.1/100	Z=-1.21 d.f.=1 P=0.2266	7.74	0.948 [ 0.87 ; 1.034 ]
rel.time.PE.uncolo	593	0/5.8/100	0/7.9/100	Z=-1.06 d.f.=1 P=0.2899	19.31	0.991 [ 0.975 ; 1.008 ]
rel.time.sasm	593	0/11.9/100	0/13/100	Z=-0.35 d.f.=1 P=0.7295	23.95	0.998 [ 0.987 ; 1.009 ]
rel.time.no.sasm	593	0/88.1/100	0/87/100	Z=0.35 d.f.=1 P=0.7295	23.95	1.002 [ 0.991 ; 1.013 ]
rel.time.PA.sasm	593	0/11.2/100	0/11.3/100	Z=0.02 d.f.=1 P=0.9863	23.02	1 [ 0.989 ; 1.011 ]
rel.time.PA.no.sasm	593	0/82.1/100	0/78.8/100	Z=0.98 d.f.=1 P=0.3258	29.71	1.005 [ 0.995 ; 1.014 ]
rel.time.PE.sasm	593	0/0.7/14.9	0/1.7/100	Z=-1.29 d.f.=1 P=0.1967	6.92	0.932 [ 0.837 ; 1.037 ]
rel.time.PE.no.sasm	593	0/6.1/100	0/8.3/100	Z=-1.08 d.f.=1 P=0.2808	19.67	0.991 [ 0.975 ; 1.007 ]
rel.time.sarm	593	0/17.3/98.1	0/15/100	Z=0.68 d.f.=1 P=0.4987	26.37	1.003 [ 0.994 ; 1.012 ]
rel.time.no.sarm	593	1.9/82.7/100	0/85/100	Z=-0.68 d.f.=1 P=0.4987	26.37	0.997 [ 0.988 ; 1.006 ]
rel.time.PA.sarm	593	0/17/97.8	0/14.2/100	Z=0.88 d.f.=1 P=0.3788	25.66	1.004 [ 0.995 ; 1.014 ]
rel.time.PA.no.sarm	593	0/76.2/100	0/75.8/100	Z=0.22 d.f.=1 P=0.8229	32.43	1.001 [ 0.993 ; 1.009 ]
rel.time.PE.sarm	593	0/0.3/6.6	0/0.8/100	Z=-0.5 d.f.=1 P=0.618	6.72	0.973 [ 0.872 ; 1.085 ]
rel.time.PE.no.sarm	593	0/6.5/100	0/9.2/100	Z=-1.22 d.f.=1 P=0.2241	20.77	0.99 [ 0.975 ; 1.006 ]
nb.sa	593	0/2.6/10	0/2.3/12	Z=1.23 d.f.=1 P=0.2191	2.14	1.083 [ 0.953 ; 1.231 ]
dens.sa	593	0/32.1/83.3	0/22.8/80	Z=4.05 d.f.=1 P=1e-04	18.73	1.031 [ 1.016 ; 1.047 ]
dur.sa	593	0/22.8/234.9	0/19.1/323.3	Z=0.68 d.f.=1 P=0.4964	44.11	1.002 [ 0.997 ; 1.007 ]
dur.PA.sa	593	0/22.5/234.8	0/18.7/322.6	Z=0.69 d.f.=1 P=0.4895	44.06	1.002 [ 0.997 ; 1.007 ]
dur.PE.sa	593	0/0.3/5.7	0/0.4/8.3	Z=-0.49 d.f.=1 P=0.6213	0.98	0.926 [ 0.681 ; 1.258 ]
nb.sarm	593	0/1.2/8	0/1.1/11	Z=0.65 d.f.=1 P=0.5133	1.41	1.066 [ 0.88 ; 1.291 ]
dens.sarm	593	0/13.6/61.5	0/11.2/60	Z=1.5 d.f.=1 P=0.1324	13.06	1.015 [ 0.995 ; 1.035 ]
dur.sarm	593	0/13.5/142	0/10.9/323.2	Z=0.66 d.f.=1 P=0.5082	32.08	1.002 [ 0.995 ; 1.009 ]
dur.PA.sarm	593	0/13.4/141.9	0/10.8/322.6	Z=0.66 d.f.=1 P=0.5084	32.04	1.002 [ 0.995 ; 1.009 ]
dur.PE.sarm	593	0/0.1/2	0/0.1/2.6	Z=0.12 d.f.=1 P=0.9021	0.31	1.057 [ 0.435 ; 2.567 ]
nb.sasm	593	0/1.4/5	0/1.3/6	Z=1.28 d.f.=1 P=0.2016	1.3	1.135 [ 0.934 ; 1.38 ]
dens.sasm	593	0/20.2/80	0/12.5/75	Z=4.13 d.f.=1 P=0	13.88	1.036 [ 1.019 ; 1.054 ]
dur.sasm	593	0/9.2/124.5	0/8.4/256.4	Z=0.31 d.f.=1 P=0.7587	24.83	1.002 [ 0.992 ; 1.011 ]
dur.PA.sasm	593	0/9/124.5	0/8.1/256.4	Z=0.33 d.f.=1 P=0.7426	24.81	1.002 [ 0.992 ; 1.011 ]
dur.PE.sasm	593	0/0.2/4.2	0/0.3/7.7	Z=-0.61 d.f.=1 P=0.5397	0.84	0.89 [ 0.612 ; 1.293 ]
prev.same.strain.nbgh	593	0/0.3/4	0/0.1/5	Z=2.36 d.f.=1 P=0.0185	0.52	1.725 [ 1.096 ; 2.716 ]
dens.same.strain	593	0/18.4/100	0/14.7/100	Z=1.63 d.f.=1 P=0.104	17.14	1.012 [ 0.998 ; 1.026 ]
nb.same.strain	593	0/1.1/5	0/1.3/6	Z=-0.58 d.f.=1 P=0.5641	1.29	0.939 [ 0.758 ; 1.163 ]
PA.same.strain	593	0/0.6/3	0/0.6/4	Z=-0.1 d.f.=1 P=0.9198	0.89	0.984 [ 0.726 ; 1.335 ]
PE.same.strain	593	0/0.5/5	0/0.6/6	Z=-0.68 d.f.=1 P=0.4956	0.94	0.902 [ 0.67 ; 1.214 ]

rel.dur.same.strain	593	0/11.9/100	0/13/100	Z=-0.35 d.f.=1 P=0.7295	23.95	0.998 [ 0.987 ; 1.009 ]	
rel.dur.PA.same.strain	593	0/11.2/100	0/11.3/100	Z=0.02 d.f.=1 P=0.9863	23.02	1 [ 0.989 ; 1.011 ]	
rel.dur.PE.same.strain	593	0/0.7/14.9	0/1.7/100	Z=-1.29 d.f.=1 P=0.1967	6.92	0.932 [ 0.837 ; 1.037 ]	
dur.cont.same.strain	593	0/9.2/124.5	0/8.4/256.4	Z=0.31 d.f.=1 P=0.7587	24.83	1.002 [ 0.992 ; 1.011 ]	
dur.cont.PA.same.strain	593	0/9/124.5	0/8.1/256.4	Z=0.33 d.f.=1 P=0.7426	24.81	1.002 [ 0.992 ; 1.011 ]	
dur.cont.PE.same.strain	593	0/0.2/4.2	0/0.3/7.7	Z=-0.61 d.f.=1 P=0.5397	0.84	0.89 [ 0.612 ; 1.293 ]	
mean.same.strain.nbgh	593	0/1.4/5.7	0/1.4/4.3	Z=-0.6 d.f.=1 P=0.551	0.92	0.899 [ 0.632 ; 1.277 ]	
mean.same.strain.nbgh.PA	593	0/0.7/2.3	0/0.8/3.7	Z=-0.72 d.f.=1 P=0.4692	0.64	0.841 [ 0.525 ; 1.345 ]	
mean.same.strain.nbgh.PE	593	0/0.6/5	0/0.7/4	Z=-0.06 d.f.=1 P=0.9546	0.71	0.989 [ 0.669 ; 1.461 ]	
mean.sarm.2hop.nbgh	593	0/1.4/5.8	0/1.4/5	Z=0.31 d.f.=1 P=0.7583	1.01	1.052 [ 0.763 ; 1.45 ]	
mean.sarm.2hop.nbgh.viaPA	593	0/0.8/4	0/0.7/4	Z=1.06 d.f.=1 P=0.2891	0.72	1.247 [ 0.829 ; 1.876 ]	
mean.sarm.2hop.nbgh.viaPE	593	0/0.6/2.7	0/0.7/4.7	Z=-0.6 d.f.=1 P=0.5487	0.72	0.885 [ 0.594 ; 1.318 ]	

### Annexe 3 : Dictionnaire des covariables utilisées aux Annexes 1 et 2

age	Âge du patient en années.
sexe	Sexe du patient.
mif	Mesure d'Indépendance Fonctionnelle (variable initialement continue recodée en quartiles).
time.at.risk	Délai en jours depuis inclusion et colonisation (cas) ou date de sélection (témoin).
atb.expo	Administration ou non d'antibiotiques au patient au cours de la semaine précédant la colonisation (cas) ou sélection (témoin).
atb.mssa.sensi.XX	Administration ou non, au cours de la semaine précédant la colonisation, d'antibiotiques dont l'efficacité (XX) contre SARM est codée S (susceptible), I (intermédiaire) ou R (résistant). ( <i>données de sensibilité de Kardaś-Słoma et al.</i> )
atb.hamrsa.sensi.XX	Administration ou non, au cours de la dernière semaine, d'antibiotiques dont l'efficacité (XX) contre les souches hospitalières de SARM est codée S (susceptible), I (intermédiaire) ou R (résistant). ( <i>données de sensibilité de Kardaś-Słoma et al.</i> )
nb.nbgh.atb.expo	Nombre de voisins du cas ou du témoin qui ont reçu des antibiotiques la semaine précédente.
nb.nbgh.atb.mssa.sensi	Nombre de voisins du cas ou du témoin qui ont reçu des antibiotiques efficaces contre SARM la semaine précédente.
nb.nbgh.atb.mrsa.sensi	Nombre de voisins du cas ou du témoin qui ont reçu des antibiotiques efficaces contre SARM la semaine précédente.
previous.sa	Historique de colonisation par <i>S. aureus</i> du cas ou du témoin (variable qualitative binomiale).
previous.sasm	Historique de colonisation par SARM du cas ou du témoin (variable qualitative binomiale).
previous.sarm	Historique de colonisation par SARM du cas ou du témoin (variable qualitative binomiale).
previous.same.strain	Historique de colonisation du cas ou de témoin par la même souche que celle actuellement isolée chez le cas (variable qualitative binomiale).
prev.XX.nbgh	Nombre de voisins du cas ou du témoin ayant eu par le passé un épisode de colonisation à XX ( <i>S. aureus</i> , SARM, SARM).
prev.XX.nbgh.YY	Nombre de voisins de type YY (tous, patient, personnel) ayant eu par le passé un épisode de colonisation à XX ( <i>S. aureus</i> , SARM, SARM).

prev.XX.nbgh.time	Durée cumulée des contacts du cas ou du témoin avec des voisins ayant eu par le passé un épisode de colonisation à XX ( <i>S. aureus</i> , SASM, SARM).
prev.XX.nbgh.YY.time	Durée cumulée des contacts du cas ou du témoin avec des voisins de type YY (tous, patient, personnel) ayant eu par le passé un épisode de colonisation à XX ( <i>S. aureus</i> , SASM, SARM).
nw.range	Taille totale, en nombre d'individus, du réseau égocentré sur le cas ou le témoin et agrégé sur les 7 jours précédant l'épisode de colonisation.
betweenness.ego	Mesure de centralité normalisée du cas ou du témoin dans le réseau agrégé sur les 7 jours précédant l'épisode de colonisation.
mean.with.lvl.1	Nombre moyen de liens des différents individus avec d'autres individus situés dans le voisinage 1-hop du cas ou du témoin.
mean.with.same.lvl	Nombre moyen de liens d'un individu du voisinage $k$ -hop du cas ou du témoin avec des individus du même voisinage $k$ -hop.
mean.at.lvl1	Nombre moyen de liens d'un individu du voisinage 1-hop du cas ou du témoin.
cpi.in.wk	Nombre de Close-Proximity Interactions (contact-jour) du cas ou du témoin au cours des 7 jours précédant la colonisation (nombre d'arêtes chaque jour qui implique le cas ou le témoin).
XX.dly.cpy	Nombre XX (maximal, moyen, médian) quotidienne de liens quotidiens impliquant le cas ou le témoin au cours des 7 jours précédant la colonisation.
XX.dur.dly.cpi	Durée XX (maximale, moyenne, médiane) quotidienne de contact impliquant le cas ou le témoin au cours des 7 jours précédant la colonisation.
degree	Nombre de voisins du cas ou du témoin au cours des 7 jours précédant la colonisation.
degree.XX	Nombre de voisins de statut XX (patient, personnel) du cas ou du témoin au cours des 7 jours précédant la colonisation.
degree.XX.YY	Nombre de voisins de statut XX (patient, personnel) du cas ou du témoin, ayant un statut de colonisation YY ( <i>S. aureus</i> , SASM, SARM) au cours des 7 jours précédant la colonisation.
degree.XXward	Nombre de voisins du cas ou du témoin au cours des 7 jours précédant la colonisation et appartenant au XX service (même, différent).

degree.XX.YYward	Nombre de voisins du cas ou du témoin, de statut XX (patient, personnel), au cours des 7 jours précédant la colonisation et appartenant au YY service (même, différent).
degree.PE.XX	Nombre de voisins du cas ou du témoin, de statut « personnel de santé », et appartenant au service XX (autre service que le cas, personnel de nuit, ergothérapeute, kinésithérapeute, autre).
nbgh.of.nbgh.bw	Centralité normalisée des contacts des voisins directs du cas ou du témoin.
nbgh.of.nbgh.nb	Nombre moyen de contacts des voisins directs du cas ou du témoin.
nbgh.of.nbgh.nb.XX	Nombre moyen de contacts de statut XX (patient, personnel) des voisins directs du cas ou du témoin.
nbgh.of.nbgh.cpi.nb	Nombre moyen de CPI impliquant les contacts des voisins directs du cas ou du témoin.
nbgh.of.nbgh.cpi.dur	Durée moyenne des CPI impliquant les contacts des voisins directs du cas ou du témoin.
nbgh.of.nbgh.cpi.dur.XX	Durée moyenne des CPI impliquant les contacts de statut XX (patient, personnel) des voisins directs du cas ou du témoin.
nbgh.of.nbgh.cpi.dly.nb	Nombre moyen de CPI chaque jour, impliquant les contacts des voisins directs du cas ou du témoin.
nbgh.of.nbgh.cpi.dly.dur	Durée moyenne des CPI chaque jour, impliquant les contacts des voisins directs du cas ou du témoin.
dens	Densité du réseau égocentré sur le cas ou le témoin, agrégé sur les 7 jours précédant la colonisation.
dens.pat	Pourcentage de patients dans le réseau égocentré sur le cas ou le témoin, agrégé sur les 7 jours précédant la colonisation.
dur.cont, dur.cont.XX	Durée moyenne de contact entre le cas ou le témoin et son voisinage 1-hop (tous, ou alors de statut XX (patient, personnel)).
dur.cont.uncolonized, dur.cont.uncolonized.XX	Durée moyenne de contact entre le cas ou le témoin et son voisinage 1-hop non colonisé par <i>S. aureus</i> au dernier prélèvement (tous, ou alors de statut XX (patient, personnel)).
dur.cont.no.sasm, dur.cont.no.sasm.XX	Durée moyenne de contact entre le cas ou le témoin et son voisinage 1-hop non colonisé par SASM au dernier prélèvement (tous, ou alors de statut XX (patient, personnel)).
dur.cont.no.sarm, dur.cont.no.sarm.XX	Durée moyenne de contact entre le cas ou le témoin et son voisinage 1-hop non colonisé par SARM au dernier prélèvement (tous, ou alors de statut XX (patient, personnel)).

rel.time.PA, rel.time.PE	Répartition du temps de contact du cas ou du témoin avec d'autres patients (PA) ou des personnels de santé (PE).
rel.time.YY, rel.time.XX.YY	Répartition du temps de contact du cas ou du témoin avec d'autres individus au statut de colonisation YY (non colonisé, <i>S. aureus</i> , SASM, non colonisé par SASM, SARM, non colonisé par SARM) et de statut XX (patient, personnel).
nb.YY, dens.YY, dur.YY, dur.XX.YY	Nombre et densité de porteurs de YY ( <i>S. aureus</i> , SASM, SARM) dans le réseau égo-centré sur le cas ou le témoin et agrégé sur les 7 jours précédant la colonisation ; et durée de contacts du cas ou du témoin avec ces individus (tous, ou de statut XX (patient, personnel)).
prev.same.strain.nbgh	Nombre de voisins du cas ou du témoin ayant eu par le passé une colonisation par la même souche que celle du cas.
dens.same.strain	Pourcentage des individus du réseau égo-centré sur le cas ou le témoin et porteurs de la même souche que celle du cas.
nb.same.strain, XX.same.strain	Nombre de voisins (éventuellement de statut XX (patient, personnel)) du cas ou du témoin, porteurs de la même souche que celle du cas.
rel.dur.same.strain, rel.dur.XX.same.strain	Répartition du temps passé par le cas ou le témoin en contact avec des porteurs de la même souche que celle du cas (éventuellement de statut XX (patient, personnel)).
dur.cont.same.strain, dur.cont.XX.same.strain	Durée cumulée passée par le cas ou le témoin en contact avec des porteurs de la même souche que celle du cas (éventuellement de statut XX (patient, personnel)).
mean.same.strain.nbgh, mean.same.strain.nbgh.XX	Nombre moyen de contacts des voisins du voisinage 1-hop du cas ou du témoin (éventuellement de statut XX (patient, personnel)), porteurs de la même souche que celle du cas.
mean.sarm.2hop.nbgh, mean.sarm.2hop.nbgh.viaXX	Nombre de porteurs de SARM parmi les contacts des voisins du voisinage 1-hop du cas ou du témoin, où l'intermédiaire du voisinage 1-hop est éventuellement de statut XX (patient, personnel).